Cariniana
Rede Brasileira de Serviços de Preservação Digital

ibict
Instituto Brasileiro de Informação em Ciência e Tecnologia

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO

GUIA DE BOAS PRÁTICAS DE PRESERVAÇÃO DE CONTEÚDO WEB

2025

DRIADE

PRESIDÊNCIA DA REPÚBLICA

Luiz Inácio Lula da Silva
Presidente da República

Geraldo José Rodrigues Alckmin Filho
Vice-Presidente da República

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

Luciana Santos
Ministra da Ciência, Tecnologia e Inovação

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA

Tiago Emmanuel Nunes Braga
Diretor

Carlos André Amaral de Freitas
Coordenação de Administração (COADM)

Ricardo Medeiros Pimenta
Coordenação de Ensino e Pesquisa em Informação para a Ciência e Tecnologia (COEPI)

Henrique Denes Hilgenberg Fernandes
Coordenação de Planejamento, Acompanhamento e Avaliação (COPAV)

Cecília Leite Oliveira
Coordenação-Geral de Informação Tecnológica e Informação para a Sociedade (CGIT)

Washington Luís Ribeiro
Coordenação-Geral de Informação Científica e Técnica (CGIC)

Alexandre Faria de Oliveira
Coordenação-Geral de Tecnologias de Informação e Informática (CGTI)
Coordenação de Governança em Tecnologias para Informação e Comunicação (COTIC)

Milton Shintaku
Coordenador de Articulação, Geração e Aplicação de Tecnologia (COTEC)

GUIA DE BOAS PRÁTICAS DE PRESERVAÇÃO DE CONTEÚDO WEB

2025

Miguel Angel Márdero Arellano
Flor de María Silvestre
(organizadores)

Autores

Rondineli Gama Saad
Danilo Formenton
Gabriela Ayres Ferreira Terrada
Márcia Ênia Lopes de Sousa
Sonia Araújo de Assis Boeres
Vanderlei Batista dos Santos

Editora Ibict

2025

Sumário

Sumário

Sumário

Sumário

APRESENTAÇÃO

01

A preservação de conteúdo digital representa um desafio e uma oportunidade. No contexto da expansão acelerada da internet, a web tornou-se um repositório dinâmico e efêmero de informações científicas, culturais, políticas e sociais. Esse caráter volátil exige esforços coordenados para garantir que o patrimônio digital seja mantido acessível e compreensível para gerações futuras. Este guia reflete o compromisso do Grupo Dríade, da Rede Cariniana, com a disseminação de boas práticas no arquivamento e na preservação digital.

A elaboração deste documento reuniu especialistas de diversas instituições, que contribuíram com suas experiências para construir um referencial prático e teórico. Este guia propõe-se a ser uma ferramenta indispensável para profissionais e organizações que buscam proteger a memória digital em um cenário de rápida transformação tecnológica. Além disso, ele ressalta a importância da preservação de conteúdo da web como um direito social, alinhado às demandas por transparência, acesso à informação e proteção da diversidade cultural na web.

Dividido em etapas claras, o guia aborda desde os critérios de seleção e tecnologias utilizadas no arquivamento, até os desafios técnicos e éticos envolvidos na manutenção de arquivos digitais. Ele explora a integração de soluções tecnológicas, como o formato WARC e ferramentas open source, com abordagens colaborativas entre instituições nacionais e internacionais. A conexão entre teoria e prática aqui apresentada destaca a relevância da construção de políticas públicas robustas que viabilizem a preservação sistemática de conteúdos digitais e fortaleçam a memória coletiva.

Por fim, este guia é mais do que um manual técnico; ele é um convite à reflexão e à ação coletiva. Reconhecendo a natureza multidisciplinar da preservação digital, o texto é voltado para bibliotecários, arquivistas, pesquisadores, gestores de políticas públicas e demais interessados. Que este material inspire iniciativas e colaborações futuras, reafirmando a importância de uma web preservada como legado histórico e instrumento de conhecimento. Assim, reafirmamos o compromisso da Rede Cariniana e do Grupo Dríade com a continuidade do saber em meio à era digital.

INTRODUÇÃO

O arquivamento de conteúdo web é o processo de captura, armazenamento e manutenção de informações digitais com o objetivo de preservar o acesso a longo prazo. Na era atual, em que o digital permeia todos os aspectos do cotidiano, desde o aprendizado até o trabalho e o lazer, a preservação desse conteúdo é fundamental para garantir que dados históricos e de interesse público estejam acessíveis no futuro (Rockembach, 2018). A web, como serviço construído sobre a Internet e composta por páginas interconectadas, depende da criação de arquivos digitais para manter acessível o que, de outra forma, se perderia ao sair do ar. Esse arquivamento envolve grandes desafios tecnológicos e organizacionais, exigindo colaboração internacional para assegurar a permanência dessa memória digital para as gerações futuras.

A crescente necessidade de preservação de conteúdo web é impulsionada pela rapidez com que as informações se tornam inacessíveis, seja pela obsolescência tecnológica, pela perda acidental ou por ataques cibernéticos. A volatilidade dos dados digitais se contrapõe à permanência da informação impressa, que possui mecanismos de preservação bem estabelecidos, como o depósito legal e os repositórios institucionais (Gomes, 2010a). Na web, no entanto, esses mecanismos tradicionais não contemplam plenamente a preservação a longo prazo, o que torna essencial a criação de estratégias e tecnologias para o arquivamento digital que garantam o acesso continuado e confiável a esses conteúdos.

A importância do arquivamento da web é inegável. Na medida em que informações de grande relevância são publicadas e rapidamente atualizadas ou removidas da web, a capacidade de reter e acessar esses registros torna-se importante. No Brasil, a falta de iniciativas consolidadas para arquivamento de sites governamentais limita o acesso prolongado a informações institucionais, evidenciando a necessidade de políticas públicas que fortaleçam a transparência e o direito ao conhecimento (Luz, 2022). O arquivamento digital, portanto, assegura que o patrimônio informacional da sociedade permaneça disponível, permitindo que a sociedade e os pesquisadores compreendam o contexto histórico e as mudanças na era digital.

Os desafios do arquivamento web são notáveis e envolvem tanto a natureza dinâmica dos sites modernos quanto a obsolescência de formatos e tecnologias. À medida que as plataformas digitais evoluem, os arquivos precisam acompanhar essas mudanças para manter a fidelidade da informação arquivada e assegurar uma experiência de uso semelhante à original (Hockx-Yu, 2012). Essa constante adaptação tecnológica é importante para que os arquivos web possam capturar com precisão conteúdos em diferentes formatos e plataformas, mas também representa uma barreira significativa, que demanda soluções inovadoras e o uso de ferramentas avançadas.

Apesar desses desafios, o arquivamento digital também apresenta oportunidades importantes. Ferramentas open sources e formatos padronizados viabilizam a captura sistemática de conteúdo web, facilitando o armazenamento de grandes volumes de dados de maneira organizada e segura (Melo, 2020). A adoção de tais tecnologias permite que instituições ajustem suas práticas de preservação digital, garantindo um acesso a longo prazo que assegure a integridade e autenticidade dos dados arquivados. Esse desenvolvimento tecnológico é um recurso valioso para enfrentar a natureza volátil da web e a diversidade de formatos digitais utilizados.

A preservação digital tem como aliada a possibilidade de automação, que a diferencia da preservação de publicações impressas, tradicionalmente dependente de processos manuais (Boeres e Saad, 2023). No entanto, ao contrário dos impressos, a informação digital precisa ser preservada rapidamente, pois o conteúdo web se torna inacessível com maior rapidez. A variedade de formatos digitais, com diferentes graus de durabilidade e compatibilidade tecnológica, amplia a complexidade da preservação, exigindo que os profissionais da área identifiquem e priorizem formatos que mantenham a integridade das informações.

Em meio a esses desafios, a colaboração internacional tem sido um componente essencial para o sucesso da preservação digital. O Consórcio Internacional para a Preservação da Internet (IIPC), formado em 2003, reúne bibliotecas, museus e arquivos de mais de 35 países, promovendo padrões e ferramentas para arquivamento de conteúdos online (Boeres e Saad, 2023). Essa iniciativa visa facilitar o desenvolvimento de uma infraestrutura global de preservação digital que suporte o acesso contínuo ao conhecimento gerado na web, permitindo que os países compartilhem práticas e aprimorem suas próprias políticas de preservação de conteúdo digital.

Em suma, o arquivamento de conteúdo web é uma prática indispensável para a manutenção da memória digital em uma era onde o conteúdo online é altamente volátil e suscetível à perda. Os desafios técnicos e organizacionais, juntamente com as oportunidades trazidas por novas tecnologias e pela colaboração entre instituições, indicam a necessidade de políticas públicas e práticas consolidadas para garantir o acesso permanente à informação digital. Com esses esforços, a preservação digital não apenas atende ao direito ao conhecimento, mas também contribui para uma sociedade mais informada e preparada para lidar com as rápidas mudanças da era digital.

ESTUDO PROSPECTIVO

08

A preservação digital de conteúdo web enfrenta desafios constantes, impulsionados pela rápida evolução das tecnologias e pelo crescimento exponencial de dados. Tecnologias emergentes, como inteligência artificial (IA), blockchain e novos formatos de armazenamento, têm potencial para transformar profundamente as estratégias de preservação digital. A IA, por exemplo, facilita a coleta e categorização de dados, identificando padrões e organizando conteúdos de maneira eficiente, enquanto o blockchain oferece um sistema descentralizado para assegurar a integridade dos arquivos. A transição gradual do formato ARC para o formato WARC ilustra essa adaptação tecnológica, promovendo interoperabilidade e permitindo a gestão de conteúdo duplicado, como salientado por Costa, Gomes e Silva (2016). Paralelamente, ferramentas como PLATO e DROID apoiam a preservação, embora sua implementação em larga escala ainda exija otimizações, conforme observa Kulovits (2009).

Essas tecnologias emergentes também contribuem para ampliar o acesso e a usabilidade dos arquivos preservados. O uso de algoritmos de IA no arquivamento da web permite a análise de grandes volumes de dados, destacando conteúdos de valor histórico, científico e cultural. Ferramentas como o Web Curator Tool (WCT), desenvolvidas para a automação do processo seletivo de arquivamento, exemplificam como a tecnologia facilita a captura de dados dinâmicos e interativos, antes dependentes de processos manuais. O Reino Unido implementou o WCT em colaboração com a Biblioteca Nacional da Nova Zelândia e a Oakleigh Consulting, um avanço que reflete a adaptação contínua do arquivamento da web aos novos formatos e fluxos de dados (Costa, Gomes e Silva, 2016).

No entanto, a inovação tecnológica não elimina a necessidade de uma análise de riscos cuidadosa para garantir a integridade e a acessibilidade dos arquivos digitais. A rápida obsolescência de formatos, somada a riscos de falhas de hardware e vulnerabilidades de segurança, apresenta ameaças à longevidade dos dados arquivados. Estudos mostram que 80% das páginas da web perdem sua forma original após um ano, ressaltando a urgência de estratégias que protejam contra perdas (Costa, Gomes e Silva, 2016). Essa vulnerabilidade, agravada pela evolução contínua da tecnologia, exige políticas de segurança digital que incluam medidas contra o desaparecimento de links e ataques cibernéticos, protegendo o valor informativo dos dados a longo prazo.

A mitigação desses riscos demanda adaptações técnicas nos mecanismos de arquivamento, especialmente em redes sociais e plataformas colaborativas, cujos conteúdos dinâmicos complicam o processo de captura e preservação. Como observado por Masanès (2006), a imprevisibilidade desses conteúdos torna o arquivamento um processo complexo e dispendioso. A necessidade de revisitar páginas para capturar atualizações regularmente consome recursos consideráveis, sem garantir a integridade total dos dados, como aponta Kulovits (2009). Para responder a esses desafios, é essencial desenvolver protocolos de arquivamento e segurança que preservem tanto a continuidade quanto o valor histórico dos conteúdos digitais.

Outro aspecto fundamental para a preservação digital é o desenvolvimento de políticas e padrões que garantam a consistência e a interoperabilidade entre diferentes sistemas e instituições. O Consórcio Internacional de Preservação da Internet (IIPC), fundado em 2003 pela Biblioteca Nacional da França, é uma iniciativa que promove diretrizes padronizadas para o arquivamento global da web, integrando atualmente mais de 45 países e instituições culturais e acadêmicas (*International Internet Preservation Consortium*, 2017). A adoção de formatos padronizados como o WARC permite que os arquivos sejam compartilhados entre instituições, criando uma rede global de preservação digital e incentivando práticas coerentes, como destacado por Costa, Gomes e Silva (2016).

Essas políticas são importantes para a seleção e organização do conteúdo, além de garantirem a autenticidade dos dados preservados. O Internet Archive, por exemplo, adota procedimentos para capturar e armazenar conteúdos críticos de relevância histórica. Durante as eleições francesas de 2002, a Bibliothèque Nationale de France implementou uma abordagem manual para arquivar conteúdos essenciais, demonstrando como políticas bem definidas asseguram a preservação de eventos culturais e políticos de valor significativo (Masanès, 2006). Esse exemplo evidencia a importância de padrões elevados de qualidade, que orientam novas iniciativas de preservação digital.

No âmbito da infraestrutura, o crescimento constante de dados arquivados exige sistemas robustos e escaláveis que garantam o armazenamento e a recuperação de dados em grande escala. Projetos como o ARQUIWEB do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), que utiliza o software Heritrix para captura de conteúdos no formato WARC, ilustram a necessidade de tecnologias compatíveis com as normas ISO para assegurar a integridade dos dados a longo prazo (IBICT, 2023). Uma infraestrutura bem planejada, composta por servidores confiáveis e dispositivos de armazenamento adequados, contribui diretamente para a estabilidade dos arquivos, permitindo que eles permaneçam acessíveis mesmo em situações adversas.

Para atender à demanda crescente, a infraestrutura de preservação digital deve ser escalável e adaptável às necessidades futuras. Muitos projetos de preservação digital buscam equilibrar automação e controle manual para manter a qualidade dos arquivos. A migração do sistema de arquivamento britânico, que passou do software PANDAS para o Web Curator Tool (WCT), exemplifica a adaptação necessária para um controle mais refinado do processo de arquivamento. Investir em infraestrutura tecnológica, portanto, não só melhora a capacidade de preservação, mas também prepara as instituições para integrar inovações que aumentam a segurança e a integridade dos dados.

Por fim, a formação de profissionais qualificados é indispensável para garantir que os processos de preservação digital sejam conduzidos conforme as políticas institucionais e atendam aos padrões de qualidade. Arquivistas e bibliotecários, em particular, devem estar capacitados para gerenciar desde a seleção de conteúdos até a aplicação de tecnologias de armazenamento, como observa Rockembach e Pavão (2018a). A criação de programas de capacitação, adaptados às necessidades de cada instituição, ajuda a implementar melhores práticas e promove uma visão estratégica voltada para a preservação digital.

Essa capacitação também é essencial para manter a qualidade e a segurança dos arquivos. O desenvolvimento de competências específicas, segundo Boeres (2017), permite que os profissionais realizem a curadoria de conteúdo e assegurem a acessibilidade dos acervos, protegendo-os de ameaças de obsolescência. Projetos como o WAX, da Harvard Library, mostram como equipes treinadas em curadoria digital podem implementar estratégias de preservação eficientes para conteúdos nato-digitais (Harvard Library, 2017). Assim, a capacitação contínua não só fortalece a eficiência das iniciativas de preservação, mas também prepara as instituições para lidar com as mudanças tecnológicas, garantindo a preservação de informações digitais valiosas.

PANORAMA NACIONAL E INTERNACIONAL DE ARQUIVAMENTO DE CONTEÚDO WEB

O arquivamento de sites institucionais é essencial para preservar informações de interesse público, garantindo que registros digitais permaneçam acessíveis ao longo do tempo. Esse processo de preservação digital compreende a coleta, armazenamento e disponibilização de informações da web, proporcionando acesso a dados importantes para futuras gerações e para o fortalecimento da memória digital de uma sociedade (Rockembach, 2018). No entanto, a situação no Brasil ainda revela uma escassez de programas estruturados para arquivar conteúdos institucionais, diferentemente do observado em países com políticas estabelecidas de preservação digital. Esse cenário reforça a importância de estabelecer práticas contínuas e robustas para o arquivamento de dados governamentais, um passo importante para fortalecer a transparência e o acesso à informação (Luz, 2022).

No cenário internacional, diversos países já avançaram em seus programas de preservação digital, implementando iniciativas sólidas para garantir o acesso contínuo ao patrimônio digital. O National Archives and Records Administration (NARA) nos Estados Unidos, por exemplo, arquiva sistematicamente páginas da Casa Branca a cada transição presidencial, protegendo o conteúdo oficial do governo (Luz, 2022). Essa prática assegura a manutenção de dados históricos importantes e serve de modelo para outras nações que desejam preservar o legado digital de suas instituições governamentais. Esse comprometimento com a preservação digital pública é uma prática que pode ser usada como referência para o desenvolvimento de políticas similares no Brasil.

A criação de políticas públicas específicas para o arquivamento de sites institucionais no Brasil encontra desafios, em grande parte pela falta de regulamentações claras e de iniciativas consolidadas. Projetos como o ARQUIWEB, do IBICT, representam esforços iniciais promissores para preservar o conteúdo digital de instituições parceiras e governamentais, buscando replicar a aparência e funcionalidade originais dos sites arquivados (Boeres e Saad, 2023). O ARQUIWEB ilustra o potencial de iniciativas nacionais em garantir o acesso continuado aos dados governamentais, mas evidencia também a necessidade de fortalecer tais programas para que atendam de maneira mais ampla à preservação da memória digital.

Internacionalmente, organizações e consórcios têm desempenhado papel importante para expandir a capacidade de arquivamento digital. Dentre eles, destacam-se o já mencionado IIPC, que une bibliotecas e instituições em mais de 35 países, facilitando o desenvolvimento e o uso de padrões e ferramentas comuns que fortalecem a preservação de conteúdo online (Boeres e Saad, 2023). Iniciativas como o Internet Archive, criado nos Estados Unidos, em 1996, e o projeto PANDORA da Biblioteca Nacional Australiana exemplificam esforços pioneiros na coleta e preservação da web, oferecendo suporte a iniciativas nacionais ao redor do mundo (Rockembach, 2018).

Esses modelos de preservação digital são importantes para países que buscam proteger seu patrimônio digital, como o Reino Unido, onde o UK Web Archive preserva sistematicamente domínios de interesse público, incluindo informações governamentais (Luz, 2022). Além disso, a experiência do Chile, que ingressou no IIPC em 2014 por meio da Biblioteca Nacional do Chile, destaca-se como um exemplo bem-sucedido na América Latina, sendo o único país da região com um programa consolidado para a preservação da web (Rockembach, 2018). Essas iniciativas ressaltam a importância da colaboração internacional e da adoção de práticas eficazes para garantir a continuidade e a acessibilidade da informação disponível na web.

No entanto, desafios significativos ainda persistem no arquivamento de conteúdos web. A obsolescência tecnológica representa um risco constante, colocando em perigo a integridade e acessibilidade dos dados digitais. Sem a implementação de políticas que garantam a integridade das informações publicadas na web muitos dos dados atualmente disponíveis podem se perder, tornando-se inacessíveis para as gerações futuras.

Além disso, o avanço tecnológico exige que as ferramentas de preservação digital sejam continuamente atualizadas para acompanhar as inovações na internet. Sites modernos frequentemente utilizam scripts dinâmicos e estruturas complexas, o que dificulta a captura completa de seus conteúdos e reduz a precisão dos registros arquivados. Hockx-Yu (2012) destaca que a qualidade da coleta é um aspecto fundamental para o sucesso no arquivamento da web, ainda que muitos desafios técnicos precisam ser superados para garantir a fidelidade dos dados armazenados.

A adoção de políticas públicas robustas é um passo importante para enfrentar esses desafios e consolidar a prática de arquivamento web no Brasil. Sem um suporte regulamentar claro, as iniciativas de preservação digital ficam fragilizadas, dependendo muitas vezes do interesse e da estrutura de organizações específicas. A falta de políticas consistentes compromete a continuidade e acessibilidade do patrimônio digital, reforçando a necessidade de regulamentações específicas para dar suporte a essa prática de longo prazo (Melo, 2020). Esse contexto coloca o Brasil em uma posição de vulnerabilidade quanto à proteção da sua memória digital.

Em meio a esses desafios, a capacitação de profissionais para atuar em preservação digital é uma oportunidade importante para expandir o alcance e a eficácia das iniciativas nacionais. Profissionais qualificados são essenciais para garantir que práticas de arquivamento digital sejam implementadas de maneira consistente, respeitando critérios de acessibilidade, conformidade e segurança da informação. Boeres e Saad (2023) destacam que aspectos como acessibilidade e coesão de metadados são fundamentais para a eficácia das práticas de arquivamento, e a capacitação de equipes pode fazer uma diferença significativa para a preservação do patrimônio digital.

Por fim, a preservação digital apresenta oportunidades de inovação e desenvolvimento de novas ferramentas para auxiliar o processo de arquivamento da web. Softwares como o Heritrix e o formato WARC têm se consolidado como soluções importantes para o armazenamento de dados digitais, permitindo que organizações ajustem suas estratégias de coleta para responder à necessidade crescente de manter o acesso a informações historicamente relevantes (Melo, 2020). O avanço na adoção dessas tecnologias pode transformar a forma como países, incluindo o Brasil, abordam a preservação digital, fortalecendo a memória digital e garantindo o acesso a dados de interesse público para futuras gerações.

ARQUIVAMENTO - PRIMEIRA ETAPA: Critérios de seleção

Para iniciar a contextualização sobre os critérios de seleção no arquivamento da web é importante entender a importância da preservação digital. O arquivamento de conteúdos publicados na internet visa a manutenção de informações relevantes que, de outra forma, poderiam se perder devido a constantes avanços tecnológicos, atualizações de plataformas e possíveis exclusões acidentais ou intencionais (Boeres e Saad, 2023). A preservação da história digital na web permite o acesso a longo prazo ao patrimônio cultural e informacional da sociedade, beneficiando áreas como a pesquisa, a educação e a transparência pública. Em um cenário em que a web está em constante expansão e transformação, a coleta e armazenamento de seus dados se tornam importantes para evitar a volatilidade inerente dos conteúdos online e garantir sua integridade para futuras gerações.

O arquivamento da web segue um fluxo contínuo de seleção e coleta que se inicia com a definição de uma política clara e abrangente. Essa política de seleção inclui a avaliação de quais conteúdos deverão ser preservados e os métodos de captura e armazenamento para garantir que a aparência e navegabilidade originais dos sites sejam mantidas (Lohndorf, 2013; Rockembach, 2019). Tal política é essencial para que os recursos coletados sejam compatíveis com os objetivos da instituição preservadora, além de oferecer uma estrutura formal para a seleção dos conteúdos mais relevantes, seja em termos de interesse cultural, científico ou social. Essa abordagem inicial é importante para a criação de um acervo que possa ser acessado de forma eficiente, reproduzindo a experiência original do usuário no momento em que o conteúdo foi arquivado.

Os critérios de seleção para o arquivamento web variam amplamente entre instituições e países, influenciados pelos objetivos e recursos disponíveis em cada contexto. Vlassenroot et al. (2021) identificam essa variação como consequência das diferentes abordagens culturais e institucionais adotadas ao redor do mundo. As diretrizes de seleção podem incluir, por exemplo, conteúdos relacionados à ciência, educação, políticas públicas ou temas de interesse social, como a desinformação e os direitos das minorias (Bibliothèque Nationale du Luxembourg, 2022). A definição clara dos temas prioritários permite que as instituições desenvolvam políticas de arquivamento que não apenas protegem a memória digital, mas que também ajudam a mitigar o risco de perda de dados valiosos em tempos de mudanças rápidas e crises informacionais.

A profundidade da coleta é outra dimensão essencial dos critérios de seleção. Existem abordagens distintas, que variam entre a coleta extensiva - que abrange uma vasta quantidade de sites em um nível superficial - e a coleta intensiva, que se aprofunda em domínios específicos, preservando conteúdos mais detalhados e camadas estruturais, como arquivos PDF e bancos de dados (Masanès, 2006). Essa distinção é relevante porque o escopo da coleta depende tanto da relevância do conteúdo quanto dos recursos técnicos e financeiros disponíveis. Em situações que demandam preservação integral, a coleta intensiva garante que toda a hierarquia do site seja mantida, enquanto, em outros casos, uma abordagem mais ampla e superficial pode atender às necessidades da instituição de forma mais eficiente.

Outro aspecto na definição dos critérios de seleção é a delimitação temática, topológica e de gênero. A escolha de coletar apenas partes de um site, como vídeos, documentos ou hyperlinks, demonstra que o arquivamento é moldado pelos limites das necessidades institucionais e dos recursos disponíveis (Rockembach e Pavão, 2018). Esse processo seletivo reflete o escopo temático da coleção, sendo essencial para otimizar o uso de recursos e assegurar a pertinência dos conteúdos arquivados. Dessa forma, o processo de seleção estabelece o que será arquivado e o que será descartado, conforme a relevância dos dados para o público-alvo e a política da organização.

A aplicação desses critérios, contudo, apresenta desafios, pois as decisões de seleção requerem um grau considerável de subjetividade. A falta de critérios claros ou a ausência de um consenso entre os gestores do projeto podem levar a decisões incoerentes ou a uma cobertura limitada de certos conteúdos (Rockembach e Pavão, 2018). Portanto, torna-se necessário estabelecer diretrizes explícitas que descrevam as áreas temáticas prioritárias e os formatos suportados, além de definir o nível de profundidade de coleta para garantir uma maior precisão na preservação. Esses critérios também servem para legitimar as escolhas feitas pelos responsáveis pelo arquivamento, reduzindo a subjetividade e promovendo uma abordagem mais estruturada e fundamentada.

As ferramentas utilizadas no processo de arquivamento, como a Wayback Machine, permitem a visualização e navegação em versões arquivadas de sites ao longo do tempo, proporcionando um acesso detalhado ao histórico de conteúdos. Esses sistemas, ao oferecerem um registro contínuo de transformações digitais, facilitam a preservação e recuperação de dados importantes. A evolução constante dessas tecnologias é vital para garantir que as informações arquivadas permaneçam acessíveis e com boa navegabilidade, aproximando-se ao máximo da experiência original do usuário.

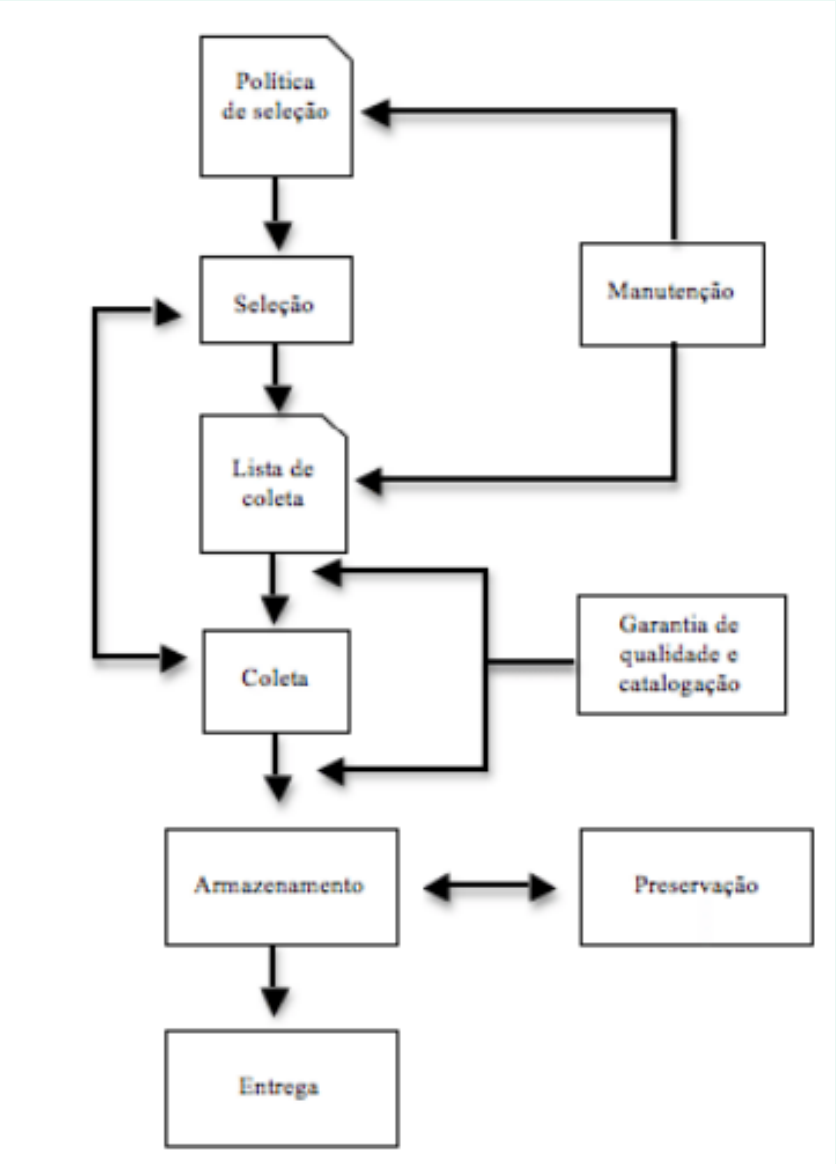
A complexidade do arquivamento da web também se reflete nos obstáculos enfrentados pelas instituições, entre eles a heterogeneidade dos conteúdos, tanto em termos de software e formatos quanto de linguagens utilizadas. As características da web, como hipertextualidade, interatividade e multimodalidade, acrescentam camadas de dificuldade ao processo de arquivamento, tornando o uso de ferramentas automatizadas essencial para lidar com a vasta quantidade de informações produzidas diariamente (Brugger e Finnemann, 2013). Ao mesmo tempo, a existência de conteúdos dinâmicos, como mídias sociais e vídeos integrados, exige que as instituições determinem a profundidade de coleta adequada para manter a integridade do conteúdo sem sobrecarregar os recursos.

Por último, a definição de critérios de seleção para o arquivamento web visa criar uma memória digital coesa, assegurando que os conteúdos preservados representem uma amostra importante e confiável da web. A adoção de diretrizes sólidas e flexíveis permite que as instituições respondam às demandas sociais e tecnológicas, preservando conteúdos de relevância histórica e cultural. Dessa forma, o arquivamento da web cumpre um papel central na construção de um legado informacional que poderá ser acessado e explorado pelas gerações futuras, evidenciando a importância da continuidade e da acessibilidade do conhecimento digital acumulado ao longo do tempo.

ARQUIVAMENTO - PRIMEIRA ETAPA: Processo de Arquivamento

O processo de arquivamento da web, conforme descrito por Brown (2006), segue uma série de etapas interconectadas para garantir a preservação e acessibilidade dos conteúdos digitais. Abaixo está uma descrição detalhada das etapas ilustradas na figura:

Figura 1: Processo de arquivamento da web



Fonte: BROWN, 2006 (tradução nossa)

- **Política de Seleção:** A política de seleção é o ponto de partida e serve como base para todo o processo. Ela define os critérios e diretrizes que orientarão quais conteúdos serão arquivados, estabelecendo prioridades de acordo com a relevância e objetivos institucionais.
- **Seleção:** Com base na política de seleção, ocorre a etapa de escolha dos conteúdos que serão arquivados. Nessa fase, são aplicados os critérios estabelecidos para determinar quais sites ou conteúdos digitais serão incluídos no processo de arquivamento.
- **Lista de Coleta:** Após a seleção, é gerada uma lista de coleta, que contém os URLs ou recursos específicos que deverão ser capturados. Essa lista é um documento dinâmico que pode ser ajustado de acordo com as mudanças nos critérios de seleção ou nas necessidades de arquivamento.
- **Coleta:** Nesta etapa, os conteúdos são efetivamente capturados por meio de ferramentas de rastreamento, conhecidas como “crawlers”, que seguem os links e armazenam o conteúdo das páginas especificadas na lista de coleta. A coleta pode ser ajustada em termos de profundidade e abrangência, dependendo dos objetivos e dos recursos disponíveis.

- **Garantia de Qualidade e Catalogação:** Após a coleta, os conteúdos passam por um processo de garantia de qualidade para verificar se foram capturados corretamente e se mantêm a integridade estrutural e funcional. Além disso, é feita a catalogação, onde os conteúdos são organizados e descritos para facilitar o acesso e a recuperação futura.
- **Manutenção:** A manutenção é uma etapa contínua que garante que a política de seleção seja atualizada conforme necessário. Ela envolve a revisão e atualização dos critérios de seleção e das listas de coleta para refletir mudanças nas necessidades institucionais ou na web.
- **Armazenamento:** Após a coleta e catalogação, os conteúdos são armazenados em repositórios seguros e duráveis, onde podem ser mantidos a longo prazo. O armazenamento é uma etapa crucial para garantir que o conteúdo esteja protegido contra perda ou corrupção.
- **Preservação:** A preservação é uma atividade complementar ao armazenamento e envolve técnicas para garantir a longevidade e acessibilidade dos conteúdos arquivados. Essa etapa pode incluir a migração de formatos e a aplicação de metadados para preservar a usabilidade futura do conteúdo.
- **Entrega:** Finalmente, os conteúdos arquivados são disponibilizados ao público ou a usuários específicos, dependendo da política de acesso da instituição. A entrega envolve a criação de interfaces que permitam aos usuários navegar pelos conteúdos arquivados de forma intuitiva e eficiente.

Esse fluxo de trabalho visa assegurar que o processo de arquivamento da web seja sistemático, confiável e alinhado com os objetivos de preservação digital, conforme descrito por Brown (2006).

ARQUIVAMENTO - PRIMEIRA ETAPA: Profundidade de coleta das páginas

Os níveis de coleta no arquivamento da web referem-se a abordagens distintas de captura de conteúdo, estabelecendo uma estrutura que permite definir a profundidade e a extensão das informações a serem preservadas. Dada a natureza efêmera da web e o crescente volume de dados, a escolha do nível de coleta é essencial para garantir que o conteúdo digital significativo seja armazenado de maneira acessível e duradoura. Segundo Masanès (2006), o arquivamento digital pode ser organizado em duas abordagens principais: extensiva e intensiva. Essas metodologias oferecem caminhos distintos para coletar e manter registros digitais, variando entre a abrangência de muitos sites de maneira superficial e o mergulho profundo em conteúdos de interesse específico.

O arquivamento extensivo, também conhecido como abordagem de amplitude, privilegia uma coleta horizontal, abrangendo uma variedade maior de sites, mas sem necessariamente registrar detalhes internos de cada um. Em contrapartida, o arquivamento intensivo busca preservar conteúdos detalhados e completos em um conjunto mais restrito de sites, priorizando a profundidade. Masanès (2006) destaca que a profundidade de captura é um fator decisivo entre essas abordagens, onde a extensão busca uma amostra ampla e superficial da web, enquanto a intensidade busca uma visão aprofundada de elementos internos, como bancos de dados e páginas hierárquicas.

O arquivamento extensivo é uma abordagem que se concentra em capturar uma visão geral da web em um dado momento. Aplicado em contextos de grande escala, como o Internet Archive, este método adota uma coleta ampla que, embora superficial, possibilita uma visão panorâmica da web. Com isso, proporciona uma amostra representativa que abrange diversas áreas temáticas e permite a análise de transformações ao longo do tempo (Masanès, 2006). Essa abordagem também facilita a observação de tendências e a formação de uma memória digital acessível, permitindo que pesquisadores acompanhem a evolução do conteúdo da web e avaliem fenômenos sociais e culturais.

Utilizar o arquivamento extensivo é particularmente vantajoso em situações onde a necessidade de cobertura é mais relevante do que a profundidade de cada site. Essa abordagem permite capturar uma ampla gama de dados, mantendo um registro básico e acessível de muitos sites e gerando um panorama de conteúdos digitais. Contudo, ao optar por uma cobertura horizontal, o método extensivo pode limitar o acesso a detalhes mais profundos, como páginas internas e interligadas de um site. Segundo Pennock (2013), o arquivamento extensivo foca em registrar conteúdos temporários de uma maneira que possibilite consultas futuras, mas sem os detalhes aprofundados que são característicos de um método mais intensivo.

Por outro lado, o arquivamento intensivo adota uma abordagem focada na profundidade, concentrando-se em registrar minuciosamente o conteúdo de poucos sites específicos. Esse método é ideal para coleções que exigem uma preservação completa, como arquivos de eventos relevantes ou temáticos, onde é fundamental manter a navegação interna e os elementos conectivos entre páginas. A abordagem intensiva permite capturar a estrutura integral de um site, incluindo todas as suas camadas e links internos, preservando o contexto completo e a experiência de navegação original. De acordo com Masanès (2006), essa abordagem é particularmente útil para registros detalhados de eventos que demandam um arquivamento profundo para uma compreensão mais rica.

O uso do arquivamento intensivo é especialmente apropriado para instituições e iniciativas que visam criar um registro histórico completo e preciso de determinados conteúdos digitais. Nesse sentido, o método intensivo assegura a captura integral dos dados essenciais, favorecendo um conjunto documental que poderá ser utilizado para consultas detalhadas e análises aprofundadas de eventos específicos ou temas de relevância histórica e científica. No entanto, é uma abordagem que demanda maior investimento de recursos, tempo e processamento. Como Pennock (2013) aponta, essa prática permite uma coleta mais rica, embora sacrifique a abrangência, limitando-se a uma seleção menor de sites.

A escolha entre o arquivamento extensivo e o intensivo envolve uma série de critérios, incluindo a relevância histórica, a acessibilidade e o custo do processo de coleta. Instituições como a Library of Congress, por exemplo, estabelecem diretrizes rigorosas para a coleta de conteúdo com base na importância cultural e científica dos sites, delimitando capturas que possam atender às necessidades de preservação sem comprometer recursos de forma excessiva (Library of Congress, 2022b). Essa análise facilita decisões sobre quais sites serão arquivados extensivamente ou intensivamente, orientando o processo de preservação digital de forma a maximizar o acesso e o valor informacional da coleção.

Há também uma abordagem híbrida que pode ser utilizada em casos específicos, combinando aspectos do arquivamento extensivo e intensivo. Esse modelo permite que instituições ajustem a profundidade e a abrangência de suas coletas conforme os objetivos de preservação e a natureza dos conteúdos. Como sugere Masanès (2006), uma abordagem balanceada entre extensão e profundidade pode ser alcançada ao aplicar a coleta extensiva em intervalos regulares, complementada por um arquivamento intensivo de eventos ou temas específicos. Isso permite um equilíbrio que otimiza o uso de recursos e atende tanto à necessidade de um panorama geral quanto à de uma captura detalhada em casos específicos.

ARQUIVAMENTO - PRIMEIRA ETAPA: Metadados

Os metadados são informações estruturadas que servem para descrever, identificar e localizar objetos digitais, desempenhando papel essencial no contexto da preservação de conteúdo web (Melo e Rockembach, 2023). Eles se classificam em várias categorias, cada uma com funções específicas, como metadados descritivos, estruturais, administrativos, técnicos e de preservação. Metadados descritivos visam facilitar a busca e recuperação dos documentos; os estruturais relacionam documentos digitais hierarquicamente, facilitando sua navegação e entendimento em contextos complexos; os administrativos ajudam na gestão dos recursos eletrônicos, incluindo detalhes sobre sua criação e manutenção; enquanto os técnicos descrevem os aspectos formais dos arquivos, como formato e resolução (Formenton e Gracioso, 2022). Já os metadados de preservação documentam as ações tomadas para garantir a integridade e acessibilidade contínua dos objetos digitais, um aspecto importante no contexto de arquivos digitais (Martins e Rockembach, 2019).

No arquivamento da web, os metadados têm uma função central, pois tornam as páginas arquivadas mais fáceis de localizar e acessar. Esses dados incluem informações essenciais, como data de arquivamento, URL, título e autor da página, elementos que ajudam a preservar a autenticidade e o contexto original do conteúdo (Venlet et al., 2018). Metadados descritivos e estruturais, por exemplo, ajudam a manter a navegabilidade e a interatividade do conteúdo original, permitindo que as páginas arquivadas sejam recuperadas com o mesmo formato e estrutura que possuíam no momento de captura (Rockembach, 2019). Esse detalhamento permite que o conteúdo digital seja contextualizado com precisão, o que facilita tanto a busca quanto a compreensão dos materiais arquivados.

Além disso, a utilização de padrões de metadados é importante para a interoperabilidade entre diferentes sistemas de arquivamento. Esse aspecto é particularmente relevante para iniciativas de arquivamento da web em escala global, onde a padronização possibilita o compartilhamento e a recuperação de conteúdos em diferentes plataformas e contextos institucionais (Formenton e Gracioso, 2022). A interoperabilidade assegurada pelos metadados é alcançada por meio de práticas e padrões de descrição que estruturam os dados de maneira uniforme, facilitando a integração entre sistemas e ampliando o alcance e o acesso ao conteúdo preservado (Masanès, 2006).

Os metadados no contexto do arquivamento da web são utilizados para registrar aspectos fundamentais de um documento digital, incluindo o contexto e as condições de captura. Esse registro detalhado permite que o conteúdo digital seja reproduzido fielmente, respeitando sua configuração original e as mudanças que ocorreram ao longo do tempo, o que garante a autenticidade e integridade do material arquivado (Rockembach, 2019). Essas informações são importantes para a preservação digital, pois possibilitam que o conteúdo continue acessível e compreensível, mesmo diante de evoluções tecnológicas que poderiam comprometer o acesso a arquivos obsoletos (Venlet et al., 2018).

O uso dos metadados também é importante para preservar o contexto e o significado original dos conteúdos arquivados, assegurando que o material digital seja interpretado corretamente pelos usuários futuros (Melo et al., 2023). Esse cuidado com o contexto histórico e funcional do conteúdo digital reflete o compromisso das instituições de preservação com a manutenção de uma experiência autêntica para o usuário, uma vez que detalhes como data de criação, localização e configuração inicial do documento podem influenciar diretamente a compreensão e o valor histórico do arquivo digital (Formenton e Gracioso, 2024).

Desenvolver práticas consistentes para a criação de metadados adaptados às características únicas dos sites e coleções arquivadas é um dos desafios enfrentados pelas iniciativas de preservação digital. Isso é especialmente relevante para metadados descritivos, que precisam representar de forma precisa e detalhada a identidade do conteúdo para assegurar sua descoberta e recuperação futura. A ausência de uma abordagem comum dificulta a criação de metadados padronizados, o que pode comprometer a integridade e a acessibilidade do conteúdo a longo prazo.

A preservação digital baseada em metadados consistentes envolve desde a coleta até a descrição detalhada do conteúdo arquivado. Metadados de preservação, por exemplo, registram a configuração original e as mudanças documentadas ao longo do tempo, assegurando que o conteúdo digital seja preservado e acessado com a mesma confiabilidade com que foi capturado inicialmente (Masanès, 2006). Esse processo possibilita que o conteúdo mantenha sua relevância e acessibilidade, contribuindo para que futuras gerações possam acessar a história digital sem distorções ou perdas significativas de informação (Martins e Rockembach, 2019).

Finalmente, o uso de metadados na organização e no acesso aos objetos digitais permite que pesquisadores e instituições explorem dados arquivados com maior precisão e eficiência (Formenton e Gracioso, 2024). Além de facilitar a pesquisa e a recuperação de informações, os metadados promovem uma organização lógica e estruturada dos conteúdos digitais, funcionando como uma espécie de índice que conecta os usuários a coleções arquivadas e permite a navegação entre documentos relacionados. Essa estrutura de acesso é importante para garantir que as coleções de conteúdo web arquivadas permaneçam relevantes e acessíveis para diversos tipos de públicos e finalidades.

ARQUIVAMENTO - PRIMEIRA ETAPA: Metadados em arquivos WARC

No contexto específico do arquivamento da web, os arquivos WARC (Web ARChive Container) não apenas armazenam os dados capturados (páginas HTML, imagens, scripts, entre outros), mas também embutem metadados essenciais em cada registro individual. Esses metadados são organizados em categorias distintas, como descritivos, técnicos, administrativos, estruturais e de preservação, cada uma com funções específicas para a gestão e o acesso sustentável do conteúdo ao longo do tempo (Melo e Rockembach, 2023; Formenton e Gracioso, 2022).

Nos arquivos WARC, os metadados descritivos permitem identificar e recuperar os recursos arquivados, registrando dados como a URL original, o título da página e a data e hora da captura. Já os metadados técnicos e administrativos documentam aspectos formais e operacionais do processo de captura, como o tipo MIME do conteúdo, o tamanho do objeto, o IP do servidor de origem, e o agente de coleta utilizado (por exemplo, Heritrix ou Browsertrix Crawler). Essas informações, contidas nos cabeçalhos dos registros WARC, são essenciais para interpretar corretamente os objetos arquivados, além de facilitar sua renderização futura (Rockembach, 2019).

Além disso, os arquivos WARC se beneficiam do uso de índices auxiliares no formato CDX, que resumem os metadados principais de cada captura e possibilitam o acesso eficiente ao conteúdo arquivado. Em termos práticos, um arquivo CDX atua como um índice cronológico que associa URLs a seus respectivos registros WARC, incluindo informações como o timestamp da captura, o código de status HTTP, o tipo de mídia, e o hash do conteúdo (digest), além da posição do registro no arquivo. Esses dados são fundamentais para viabilizar a navegação temporal em ferramentas como o Wayback Machine e o ReplayWeb.page, permitindo recuperar versões anteriores de uma página de forma precisa e contextualizada (Maemura et al., 2018).

A integração de metadados de preservação, como hashes criptográficos e logs de eventos, complementa esse ecossistema, garantindo que cada objeto digital possa ser validado quanto à sua integridade e autenticidade. Isso é especialmente relevante em contextos de reuso acadêmico e transparência institucional, onde a confiabilidade das fontes arquivadas deve ser mantida. A adoção de padrões amplamente aceitos — como o próprio WARC (ISO 28500:2017), o formato CDX e esquemas como PREMIS, METS ou Dublin Core — assegura a interoperabilidade entre sistemas de arquivamento e repositórios digitais, promovendo o compartilhamento e a reutilização de conteúdos preservados em escala global (Masanès, 2006; Martins e Rockembach, 2019).

A tabela a seguir resume os principais tipos de metadados presentes em arquivos WARC e suas respectivas funções na preservação digital:

Tabela 1: Tipos de metadados em arquivos WARC

Tipo de Metadado	Descrição	Exemplo no WARC/CDX	Função na Preservação Digital
Descritivo	Facilita a identificação e descoberta do conteúdo arquivado	URL original, título da página, tipo de recurso (HTML, imagem etc.), linguagem do conteúdo.	Apoia a indexação e recuperação da informação por humanos e sistemas de busca.
Estrutural	Relaciona objetos digitais e define a hierarquia entre eles.	Referência entre página HTML e recursos embutidos (CSS, JS, imagens); sequência de navegação.	Mantém a navegabilidade e a relação entre os componentes do site arquivado.
Técnico	Informa características formais dos objetos digitais capturados.	Tipo MIME, tamanho do objeto, codificação de caracteres, compressão (gzip).	Garante que os objetos possam ser processados, renderizados e preservados corretamente em diferentes ambientes.

Administrativo	Descreve aspectos de gestão e controle dos objetos.	Data/hora da captura (timestamp), agente de coleta (user-agent), IP do servidor, UUID do registro.	Permite rastrear a origem da captura, controlar versões e monitorar alterações nos objetos arquivados.
De Preservação	Documenta ações e condições relacionadas à preservação de longo prazo.	Hash (SHA1) do conteúdo, políticas de retenção, logs de auditoria, versionamento, eventos PREMIS.	Assegura a autenticidade, integridade e proveniência do objeto digital ao longo do tempo.
De Índice (via CDX)	Resume e organiza os registros WARC para facilitar o acesso eficiente.	CDX: URL canônica, data da captura, digest (hash), offset e comprimento no WARC, código HTTP, tipo MIME.	Viabiliza navegação temporal, recuperação rápida de registros e integração com ferramentas como o Wayback Machine.

Fonte: Elaborado pelos autores

A riqueza e a padronização dos metadados embutidos nos arquivos WARC são condições para o sucesso de iniciativas de arquivamento da web. Esses metadados não apenas promovem o acesso, a interoperabilidade e a preservação de longo prazo, mas também garantem a confiabilidade e a rastreabilidade dos conteúdos arquivados, consolidando-se como um dos pilares da preservação digital na era da informação.

ARQUIVAMENTO - PRIMEIRA ETAPA: Tecnologias e formatos de coleta

Ao longo dos anos, surgiram diversas ferramentas para o arquivamento da web, refletindo avanços importantes na preservação digital e no acesso a conteúdos online. Esse tipo de arquivamento visa capturar, armazenar e preservar informações que estariam sujeitas ao desaparecimento devido à natureza volátil da internet. A necessidade de garantir o acesso contínuo a informações históricas, culturais e acadêmicas ao longo do tempo impulsionou o uso de formatos como o WARC (Web ARChive Container), que se tornou um padrão ISO em 2009, adotado amplamente na comunidade de arquivamento da web. Como explica Masanès (2006), os arquivos WARC possibilitam a preservação detalhada de diversos aspectos do conteúdo digital, fornecendo registros e metadados técnicos que facilitam tanto a navegação temporal quanto a análise de dados.

A Internet Archive, por exemplo, criadora do Heritrix, contribui para o arquivamento em escala global, oferecendo a bibliotecas e arquivos uma solução robusta para preservar sites e conteúdos digitais. O Heritrix foi desenvolvido em Java e é altamente configurável, com opções que vão desde autenticação e filtros de conteúdo até o agendamento de tarefas. Essa flexibilidade o torna ideal para instituições de preservação digital, que podem adaptá-lo conforme suas necessidades, seja para capturar grandes quantidades de dados ou para realizar coletas específicas e detalhadas (ROCKEMBACH; PAVÃO, 2024).

Na coleta de conteúdos digitais, ferramentas como o Archive-It desempenham papel central, oferecendo flexibilidade e adaptabilidade a diferentes contextos institucionais. O Archive-It é amplamente utilizado por instituições acadêmicas e de preservação cultural, permitindo capturar e catalogar coleções de conteúdo, que podem ser armazenadas em formato WARC e acessadas a partir dos repositórios do Internet Archive. Essa ferramenta facilita a visualização dos conteúdos arquivados através do Wayback Machine, oferecendo uma busca prática por texto completo e URLs (FORMENTON, 2023).

Tabela 2: Comparação entre Heritrix e Archive-It

Característica	Heritrix	Archive-It
Tipo de uso	Ferramenta open-source para coleta autônoma	Ferramenta de código proprietário e serviço baseado em nuvem
Usuários-alvo	Equipe técnicas e instituições avançadas	Bibliotecas, arquivos e instituições públicas
Configuração e controle	Alto (requer conhecimento técnico)	Médio (interface amigável)
Escalabilidade	Alta (controle completo da infraestrutura)	Alta (infraestrutura gerenciada)
Interface gráfica	Sim	Sim
Visualização integrada	Não (necessita ReplayWeb.page ou similar)	Sim (via Wayback Machine)
Agendamento de coletas	Sim	Sim
Suporte a JavaScript	Limitado	Parcial
Formato de saída	WARC	WARC
Custo	Gratuito e open-source	Pago (por volume de dados coletados)

Fonte: Elaborado pelos autores

Quando falamos de ferramentas de captura personalizada podemos falar sobre o Conifer. O Conifer é uma ferramenta popular para capturar sessões de navegação específicas. Desenvolvido pela Rhizome, ele possui uma interface amigável e permite a criação de arquivos WARC a partir da navegação gravada, preservando conteúdos que podem escapar dos rastreadores convencionais (MASANÈS, 2023). Esse recurso é particularmente útil para pesquisadores e arquivistas que, sem conhecimentos técnicos avançados, necessitam de uma solução eficiente e personalizada para capturar conteúdos da web.

Por sua vez, o Webrecorder - ecossistema de software livre no qual o Conifer se baseia - oferece uma arquitetura mais ampla e flexível para a captura e reprodução de páginas web interativas. Além da gravação manual por navegadores reais, o Webrecorder inclui ferramentas como o Browsertrix Crawler (voltado para automação de capturas com suporte a JavaScript) e o ReplayWeb.page (visualização local de arquivos WARC). Dessa forma, ele atende tanto a usuários individuais quanto a instituições com fluxos de arquivamento mais escaláveis.

O diferencial dessas ferramentas personalizadas está na capacidade de capturar experiências de navegação ricas, que envolvem interações do usuário, carregamentos dinâmicos de conteúdo e respostas condicionadas por scripts. Por exemplo, conteúdos provenientes de redes sociais, sistemas de comentários ou páginas com autenticação são difíceis de serem arquivados com crawlers tradicionais como o Heritrix. O uso do Webrecorder e do Conifer nesses casos permite contornar limitações técnicas por meio da simulação da navegação humana, gravando precisamente o que é carregado na tela.

Além disso, o Webrecorder oferece suporte à preservação de sessões autenticadas e ao arquivamento de páginas com múltiplas camadas de carregamento, algo cada vez mais comum em sites modernos. Isso o torna essencial em contextos de preservação de memória institucional, cobertura de eventos em tempo real e arquivamento de plataformas complexas. Sua integração com ferramentas de visualização como o ReplayWeb.page garante não apenas a captura, mas também o acesso eficiente e fidedigno ao conteúdo preservado.

Essas características fazem do ecossistema Webrecorder uma boa alternativa e acessível para pesquisadores, jornalistas, ativistas e instituições de pequeno e médio porte que demandam agilidade, precisão e adaptabilidade na captura de conteúdo digital.

Tabela 3 : Comparação entre as Ferramentas de captura personalizada

Característica	Webrecorder	Conifer
Tipo de captura	Manual e automatizada (Browsertrix)	Manual (sessões interativas)
Suporte a JavaScript	Sim	Sim
Gravação de sessões autenticadas	Sim	Sim
Recurso de replay local	Sim (ReplayWeb.page)	Sim (ReplayWeb.page via integração)
Interface amigável	Moderada	Alta
Nível de controle de captura	Alto	Médio
Indicado para	Arquivos institucionais e pesquisadores técnicos	Arquivistas e pesquisadores não técnicos
Código aberto	Sim	Baseado em software aberto (Webrecorder)
Formato de saída	WARC	WARC

Fonte: Elaborada pelos autores

Em uma análise comparativa, embora o Heritrix e o Archive-It sejam mais adequados para coletas automatizadas em larga escala, ferramentas como o Webrecorder e o Conifer destacam-se na captura personalizada de sessões de navegação. A escolha entre essas soluções depende de diversos fatores: o volume de dados, a dinamicidade do conteúdo, os recursos institucionais e os objetivos da coleta. O Heritrix permite uma coleta seletiva baseada em filtros e agendamento, enquanto o Archive-It facilita a organização e o acesso institucional por meio de interface web. Já o Webrecorder proporciona maior fidelidade para conteúdos interativos, como aplicações baseadas em JavaScript, e o Conifer é ideal para capturas pontuais com pouco conhecimento técnico.

Uma das questões técnicas mais relevantes na preservação digital via arquivamento web é a padronização dos dados coletados, tanto para garantir sua longevidade quanto para viabilizar a interoperabilidade entre ferramentas. O uso do formato WARC (ISO 28500) é um exemplo dessa padronização, pois permite a encapsulação não apenas do conteúdo HTML e arquivos associados, mas também dos metadados da sessão de captura, incluindo timestamps, cabeçalhos HTTP e dados de contexto.

Como ressaltam Maemura et al. (2018), os arquivos WARC oferecem um padrão aceito internacionalmente, facilitando o compartilhamento de conteúdos entre plataformas, bem como a integração com sistemas de preservação digital mais amplos, como repositórios institucionais e redes cooperativas. Além disso, o suporte a formatos derivados, como o CDX — um tipo de índice estruturado — é fundamental para viabilizar a navegação eficiente dentro de grandes volumes de dados arquivados. Na prática, os arquivos CDX funcionam como tabelas de referência que registram metadados essenciais de cada recurso capturado, como URLs, datas de captura e localizações dentro do arquivo WARC. Esses índices permitem que ferramentas de reprodução, como a Wayback Machine, localizem rapidamente o conteúdo desejado, viabilizando a navegação temporal e o acesso seletivo a versões específicas de páginas web arquivadas.

Outro aspecto técnico é a visualização dos arquivos arquivados. Ferramentas como o ReplayWeb.page possibilitam a renderização local e autônoma de arquivos WARC diretamente em navegadores modernos, dispensando servidores intermediários. Isso amplia as possibilidades de uso em ambientes distribuídos, como projetos colaborativos, arquivos pessoais e instituições de pequeno porte.

Ademais, a captura de conteúdos dinâmicos impõe desafios contínuos relacionados à execução de scripts, dependência de recursos externos, carregamento assíncrono e interações do usuário. Diferentemente das páginas estáticas baseadas apenas em HTML e CSS, muitos sites contemporâneos dependem fortemente de JavaScript para gerar ou modificar seu conteúdo, exigindo que a ferramenta de captura não apenas acesse o código-fonte, mas também simule a renderização completa no navegador. Esses conteúdos frequentemente são carregados de forma assíncrona ou a partir de múltiplos domínios, exigindo respostas técnicas mais sofisticadas para garantir a completude e fidelidade do material arquivado. Tais aspectos requerem soluções capazes de replicar, com precisão, o comportamento de um navegador moderno e as ações do usuário.

Nesse contexto, ferramentas como o Browsertrix Crawler, integrante do ecossistema Webrecorder, representam um avanço significativo. Essa ferramenta adota navegadores headless baseados em Chromium para executar scripts, renderizar DOMs (Document Object Model) dinâmicos e simular interações como cliques, rolagem e autenticação. Com isso, é possível capturar páginas web em seu estado visual e funcional completo, inclusive em situações nas quais o conteúdo é revelado apenas por meio de ações do usuário ou chamadas JavaScript complexas. Tal abordagem é especialmente relevante para o arquivamento de redes sociais, páginas governamentais interativas, plataformas de streaming e outros ambientes digitais cuja estrutura é altamente responsiva e orientada a eventos.

Por fim, destaca-se a importância do versionamento, autenticação, compressão e verificação de integridade dos arquivos arquivados. Estratégias como o controle de versões, a geração de somas de verificação (hashing) e a utilização de formatos normalizados como o WARC são fundamentais para garantir a autenticidade e confiabilidade dos dados preservados, especialmente em contextos que envolvem acesso público, reuso acadêmico ou responsabilidade institucional. A adoção de tais práticas fortalece a confiabilidade da preservação digital e contribui para a rastreabilidade e validação dos objetos arquivados ao longo do tempo.

PRESERVAÇÃO - SEGUNDA ETAPA: Plano de Preservação

A instituição que planeja implementar a preservação digital necessita, inicialmente, publicizar essa intenção com a publicação de sua política de preservação digital. Essa política estabelece seu compromisso em relação a uma preservação embasada em normas e padrões aceitos pela comunidade especializada.

A política de preservação digital se desdobra em planos de preservação digital. Cada plano de preservação digital objetiva tratar acervos digitais diversos, tais como: documento filmográfico; documento fotográfico; documento sonoro; documento textual.

Os planos de preservação terão pontos comuns (por exemplo, padrão mínimo de metadados), bem como tratarão das diferenças apresentadas pelos diferentes gêneros documentais citados. O gênero documental implicará, por exemplo, na escolha de um formato de preservação adequado ao gênero em questão. Assim, um documento textual poderá ser preservado em PDF/A, enquanto um documento fotográfico poderá ser preservado no formato TIFF.

Além disso, os planos de preservação definirão outros detalhes: nome único do documento digital, inclusão de outros metadados além dos mínimos; suporte de preservação, tanto off-line quanto on-line; tamanho dos pacotes de informação de submissão (SIP); periodicidade de transferência para o arquivo da instituição; uso de assinatura digital; atendimento à Lei Geral de Proteção de Dados (Lei nº 13.709, de 14 de agosto de 2018); atendimento à Lei de Acesso à Informação (Lei nº 12.527, de 18 de novembro de 2011); dentre outras necessidades inerentes à organização.

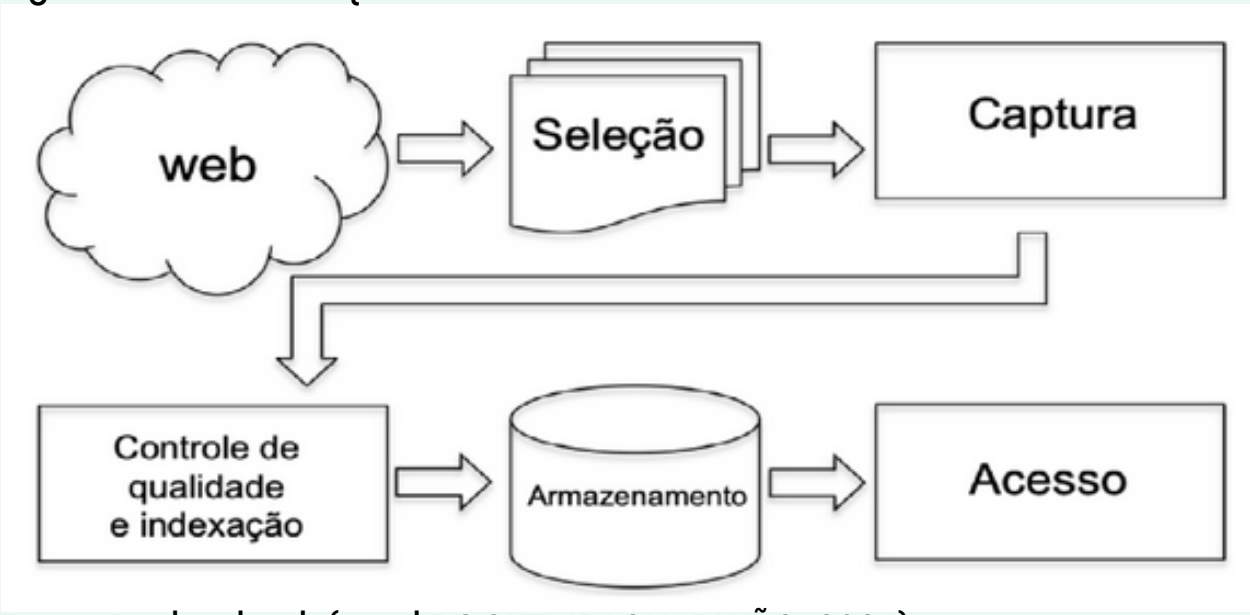
No que respeita à preservação de conteúdo web, para Márdero Arellano (apud ROCKEMBACH; PAVÃO, 2024), o plano de preservação digital aborda a preservação no longo prazo; possibilita o acesso contínuo ao acervo; fornece a documentação necessária para garantir a usabilidade; define níveis de preservação o tratamento das mudanças tecnológicas e dos requisitos de usuário.

Rockembach e Pavão (2024) apresentam algumas orientações sobre a parte geral dos planos de preservação:

1. definir o responsável pelo suporte à preservação da página web da instituição;
2. realizar o diagnóstico das condições de "arquivabilidade" das páginas da web com o uso de métodos e ferramentas aprovados pela comunidade de arquivamento da web;
3. realizar as capturas dos sites em conformidade à estrutura organizacional da instituição;
4. armazenar os sites no formato ISO WARC (web ARChive);
5. utilizar plataformas do tipo open source para o acesso aos sites preservados;
6. preservar os sites por tempo indeterminado, com a garantia de versionamento, histórico, uso e reuso dos dados;
7. analisar cada caso de solicitação de remoção de conteúdos (takedown);
8. revisar periodicamente os planos de preservação das páginas web da instituição.

Após essa introdução, os planos de preservação seguirão fluxos de trabalho, conforme segue:

Figura: 2. Fluxo de arquivamento da web



Fonte: Rockembach (apud ROCKEMBACH; PAVÃO, 2024)

Conforme Rockembach e Pavão (2024), o fluxo de arquivamento da web se constitui de etapas e atividades diferentes, que podem ser assim definidas:

1. identificar e selecionar materiais para a preservação, com estabelecimento de critérios para a determinação do que será preservado;
2. capturar os materiais com o uso de ferramentas e softwares que permitam o armazenamento de cópias completas;
3. usar técnicas de preservação digital que mantenham esses materiais por prazos indefinidos;
4. disponibilizar os materiais preservados para acesso.

O modelo Open Archival Information System (OAIS) é amplamente reconhecido como o principal referencial conceitual para sistemas de preservação digital de longo prazo. Formalizado como a norma ISO 14721:2012 – Space data and information transfer systems — Open archival information system (OAIS) — Reference model, esse modelo foi traduzido para o contexto brasileiro como a ABNT NBR 14721:2021 – Sistemas espaciais de transferência e de informação – Sistema Aberto de Arquivamento de Informação (SAAI) – Modelo de referência.

O modelo OAIS descreve uma estrutura funcional e informacional que orienta como repositórios digitais devem organizar, armazenar e fornecer acesso a objetos digitais ao longo do tempo. Um dos principais componentes dessa estrutura é a definição dos pacotes de informação, que representam diferentes estágios do ciclo de vida de um objeto digital no repositório:

- Submission Information Package (SIP): é o pacote submetido ao repositório por um produtor de conteúdo. Contém os objetos digitais e os metadados necessários para sua ingestão. O SIP pode ser originado de diversas fontes, incluindo coletores web, sistemas de produção institucional ou digitalizações, e deve conter metadados suficientes para permitir sua curadoria, identificação e transformação em AIP.
- Archival Information Package (AIP): é o pacote armazenado permanentemente pelo repositório. Resulta da transformação do SIP e incorpora o objeto digital junto a metadados estruturais, descritivos, técnicos e de preservação. O AIP é projetado para garantir a autenticidade, integridade e acessibilidade do conteúdo ao longo do tempo. Ele é a unidade básica de preservação, podendo incluir múltiplas versões ou eventos de preservação registrados no tempo.

- Dissemination Information Package (DIP): é o pacote preparado para ser entregue ao usuário final durante o processo de acesso. O DIP pode ser derivado parcial ou integralmente de um ou mais AIPs, e sua composição pode variar conforme a requisição de disseminação, política de acesso ou formato preferencial do usuário. Ele pode incluir representações simplificadas, metadados reduzidos ou transformações do conteúdo original.

Esses três tipos de pacotes (SIP, AIP e DIP) refletem as interações entre produtores, repositórios e consumidores de informação dentro do modelo OAIS, assegurando uma gestão coerente, documentada e auditável dos objetos digitais ao longo de seu ciclo de vida arquivístico.

Tabela 4 : Comparação entre os pacotes de informação no modelo OAIS

Tipo de Pacote	Função Principal	Conteúdo	Quem envia/recebe	Exemplo prático (contexto WARC)
SIP (Submission Information Package)	Submissão do objeto ao repositório	Objetos digitais originais + metadados fornecidos pelo produtor	Enviado pelo produtor para o repositório	Um conjunto de arquivos WARC gerados por uma coleta web, acompanhado de metadados descritivos (ex: título da coleção, data da coleta, escopo da captura) enviados por uma biblioteca nacional
AIP (Archival Information Package)	Preservação a longo prazo	Objeto digital + metadados de preservação, técnicos, estruturais e de proveniência	Armazenado e gerenciado pelo repositório	O mesmo WARC, enriquecido com metadados PREMIS, logs de validação, checksums, eventos de curadoria, versão de software usado, e histórico de ingestão

DIP (Dissemination Information Package)	Fornecimento de acesso ao conteúdo preservado	Conteúdo derivado + metadados de acesso, conforme políticas de disseminação	**Entregue pelo repositório ao usuário final**	**Arquivo WARC (ou um item extraído dele) fornecido ao pesquisador em formato acessível, junto com dados contextuais e restrições de uso (ex: JSON com informações do CDX index ou visualização via PyWb)**

Fonte: Elaborada pelos autores

Essa estrutura mostra como os objetos digitais (como arquivos WARC oriundos de coletas da web) transitam e se transformam dentro do repositório ao longo do tempo, com camadas sucessivas de metadados que garantem a preservação, a autenticidade e o acesso significativo.

PRESERVAÇÃO - SEGUNDA ETAPA: Empacotamento dos arquivos WARC e submissão em repositório

Recomenda-se baixar os arquivos e criar o SIP. Utilizar padrão para elaboração de planilha de metadados em .csv (comma-separated values, original para valores separados por vírgula), onde constem os metadados Dublin Core e outros metadados julgados pertinentes. A planilha assim configurada permite a interoperabilidade com a plataforma de difusão e acesso.

Deve-se utilizar software livre de preservação digital apoiado por ampla comunidade de desenvolvedores, por exemplo: Archivematica ou RODA (Repositório de Objetos Digitais Autênticos).

A plataforma de preservação digital poderá estar integrada com uma plataforma de difusão e acesso, por exemplo: AtoM ou DSpace. Essa integração diminui a quantidade de etapas necessárias para preservar e difundir o arquivo web.

PRESERVAÇÃO - SEGUNDA ETAPA:
Definição de metadados de preservação a serem incluídos no repositório

Na preservação digital, os metadados são essenciais para garantir a autenticidade, integridade e usabilidade dos documentos ao longo do tempo. Segundo Rockembach e Pavão (2024), os metadados de preservação devem descrever tanto as características dos objetos digitais quanto o histórico de intervenções realizadas para sua conservação. Esses elementos possibilitam que futuras gerações possam entender, acessar e reutilizar o conteúdo de maneira fiel ao documento original.

Os principais tipos de metadados de preservação digital incluem:

- **Metadados Descritivos:** Utilizados para identificar e descrever o conteúdo dos arquivos, os metadados descritivos fornecem informações como título, autor, data de criação e descrição do conteúdo. Isso facilita a localização e recuperação dos arquivos, mesmo após várias etapas de preservação (Rockembach e Pavão, 2024).
- **Metadados Técnicos:** Esses metadados documentam as características técnicas do arquivo digital, incluindo formato, software utilizado, resolução e outras especificidades que impactam a preservação de longo prazo. A descrição detalhada das características técnicas permite que futuras migrações de formato preservam as qualidades originais do documento, conforme indicado por Arellano (2004).

- **Metadados de Proveniência:** A proveniência trata da origem do documento e seu histórico de modificações, permitindo a rastreabilidade do objeto digital desde sua criação até o momento atual. Sayão (2010) sugere que a documentação de eventos de preservação, como migrações de formato e verificações de integridade, fortalece a autenticidade dos materiais preservados.
- **Metadados de Direitos:** Metadados de direitos incluem informações sobre restrições de acesso, direitos autorais e outras políticas de uso, conforme a Lei Geral de Proteção de Dados (Lei nº 13.709, de 2018) e a Lei de Acesso à Informação (Lei nº 12.527, de 2011). Esses metadados são fundamentais para assegurar que o conteúdo preservado seja acessado e utilizado de forma ética e conforme os regulamentos vigentes.
- **Metadados de Preservação:** O modelo PREMIS (Preservation Metadata: Implementation Strategies) é amplamente recomendado para detalhar eventos de preservação, agentes responsáveis, e procedimentos específicos realizados no objeto digital (Arellano, 2004). Esses metadados asseguram que todas as intervenções realizadas no arquivo digital sejam documentadas, incluindo as operações de verificação de integridade e armazenamento seguro.

Tabela 5: Comparação entre dos tipos de metadados na preservação digital

Tipo de Metadado	Função Principal	Exemplos Comuns	Padrões / Referências
Descritivo	Facilitar a identificação, busca e recuperação dos arquivos.	Título, autor, data de criação, assunto, palavras-chave	Dublin Core, MODS
Técnico	Documentar características técnicas para suportar a preservação de longo prazo.	Formato, tamanho, resolução, software necessário, checksums	NISO Z39.87, MIX, FITS
Proveniência	Registrar a origem e o histórico de modificações do objeto digital.	Data de ingestão, responsáveis, migrações de formato, logs	PREMIS (seção de proveniência), METS
Direitos	Informar restrições legais e políticas de acesso e uso.	Direitos autorais, licenças, nível de acesso, termos de uso	PREMIS (seção de direitos), Creative Commons
Preservação	Documentar ações de preservação, agentes envolvidos e eventos críticos.	Evento de fixidade, auditoria, backup, replicação geográfica	PREMIS, OAIS

Fonte: Elaborada pelo autor

A escolha e padronização desses metadados garantem a interoperabilidade e longevidade dos objetos digitais preservados, permitindo que diversos repositórios e plataformas possam interpretar e compartilhar essas informações.

PRESERVAÇÃO - SEGUNDA ETAPA:
Envio dos pacotes de arquivamento para um sistema de preservação

Para garantir a integridade e o acesso contínuo aos conteúdos digitais, o envio dos pacotes de arquivamento deve ser realizado de maneira metódica e embasada em padrões internacionais de preservação. Conforme o Modelo de Referência OAIS (ISO 14721:2012), o processo de envio e armazenamento no repositório requer a elaboração de pacotes específicos que assegurem a preservação a longo prazo, respeitando a integridade dos dados e permitindo futuras recuperações (Rockembach e Pavão, 2024).

Os principais passos envolvidos no envio dos pacotes de arquivamento incluem:

- **Preparação dos Pacotes de Informação de Submissão (SIP):** Cada pacote de arquivamento, também conhecido como SIP (Submission Information Package), deve conter todos os elementos necessários para a preservação, como o arquivo digital em si, a descrição completa em metadados e a documentação dos processos de preservação realizados até o momento. A organização em SIP facilita a verificação de integridade e permite a interoperabilidade com outros sistemas (Arellano, 2004).

- **Formatos e Estruturas de Arquivo para Preservação:** Recomenda-se que os arquivos sejam formatados em WARC (Web ARChive), um padrão ISO amplamente aceito na preservação de conteúdo web, conforme Rockembach e Pavão (2024). Este formato facilita o armazenamento de grandes volumes de dados e a manutenção do histórico de versões, essenciais para a longevidade do conteúdo.
- **Utilização de Softwares para Preservação Digital:** O uso de softwares open source, como Archivematica e RODA (Repositório de Objetos Digitais Autênticos), é recomendado para o preparo e envio dos pacotes de preservação. Essas plataformas suportam uma gama de funcionalidades, desde a criação e verificação dos pacotes de arquivamento até a organização dos metadados, garantindo que o processo de preservação seja realizado de forma consistente e segura (Sayão, 2010).
- **Integração com Plataformas de Acesso e Difusão:** Em alguns casos, a integração do sistema de preservação com plataformas de difusão, como DSpace e AtoM, permite que os pacotes enviados sejam prontamente disponibilizados para acesso, melhorando a usabilidade dos dados preservados. Essa integração também facilita o gerenciamento contínuo e reduz a duplicação de etapas de preservação e difusão (Rockembach e Pavão, 2024).

- **Monitoramento e Verificação Pós-Envio:** Após o envio dos pacotes para o repositório de preservação, é fundamental realizar verificações periódicas de integridade e conformidade. Essa prática, recomendada pelo Modelo OAIS, garante que os arquivos permaneçam intactos ao longo do tempo, prevenindo danos e permitindo o resgate eficiente dos dados em caso de falhas ou obsolescência tecnológica (Arellano, 2004).

Esse processo estruturado assegura que cada etapa da preservação digital seja documentada e que os pacotes de arquivamento estejam prontos para suportar futuras recuperações e uso contínuo, contribuindo para uma preservação digital confiável e duradoura.

PRESERVAÇÃO - SEGUNDA ETAPA: Recuperação sob demanda

A recuperação sob demanda é um elemento essencial em qualquer sistema de preservação digital, pois permite o acesso imediato a conteúdos específicos quando necessário, garantindo a usabilidade e acessibilidade dos dados preservados. Conforme os princípios do Modelo OAIS (ISO 14721:2012), a recuperação sob demanda envolve a disponibilização de pacotes de informação específicos, como o DIP (*Dissemination Information Package*), que é estruturado para facilitar o acesso a objetos digitais preservados (Rockembach e Pavão, 2024).

Para implementar uma recuperação sob demanda eficiente, os seguintes aspectos devem ser considerados:

- **Indexação e Organização dos Metadados:** A estruturação precisa dos metadados é fundamental para facilitar a recuperação sob demanda. Metadados descritivos, técnicos e de preservação devem estar devidamente indexados para permitir a busca eficiente e a identificação rápida dos conteúdos solicitados (Sayão, 2010). A adoção de padrões como Dublin Core e PREMIS contribui para a interoperabilidade e facilita o acesso aos arquivos.

- **Sistemas de Busca e Acesso:** A recuperação sob demanda depende de sistemas robustos de busca e acesso, que permitam consultas refinadas e acesso direto aos pacotes digitais. Plataformas como DSpace e AtoM podem ser integradas ao repositório de preservação para oferecer uma interface amigável de acesso aos documentos, permitindo a visualização e o download conforme as permissões definidas pela instituição (Arellano, 2004).
- **Controle de Acesso e Políticas de Direitos:** É necessário definir políticas de direitos que regulem o acesso aos materiais preservados, atendendo à Lei Geral de Proteção de Dados (Lei nº 13.709, de 2018) e à Lei de Acesso à Informação (Lei nº 12.527, de 2011). Essas políticas asseguram que a recuperação dos conteúdos respeite a privacidade e os direitos autorais, estabelecendo quem pode acessar quais conteúdos e em quais condições (Rockembach e Pavão, 2024).
- **Monitoramento da Integridade e Funcionalidade:** Para garantir a confiabilidade do conteúdo recuperado, recomenda-se o uso de verificações periódicas de integridade, que atestam que os arquivos preservados mantêm suas características originais. Ferramentas de verificação de integridade, como checksums, ajudam a identificar alterações ou corrupções nos arquivos e a garantir que os dados recuperados sejam idênticos aos originais (Sayão, 2010).

- **Assistência Técnica e Suporte para Usuários:** Oferecer suporte técnico para usuários que acessam os materiais sob demanda é importante para assegurar uma experiência de recuperação eficiente. Esse suporte pode incluir tutoriais, guias de acesso e atendimento para resolver dúvidas ou problemas técnicos durante o processo de recuperação.

A implementação de uma recuperação sob demanda eficaz é um diferencial em sistemas de preservação digital, pois permite o uso contínuo e controlado dos dados, garantindo que o conteúdo digital preservado esteja disponível para consultas e reutilizações conforme necessário, sempre em alinhamento com as diretrizes de preservação e políticas institucionais.

LEGISLAÇÃO - TERCEIRA ETAPA

A preservação de *websites*, embora tenha característica fundamentalmente tecnológica, pode ser feita sob diversos vieses, se interpretada no escopo de algumas disciplinas científicas, quais sejam: arquivologia, biblioteconomia e museologia. Essa abordagem parte do entendimento do que é um *website*. Se entendido como um documento produzido por uma instituição no exercício de suas funções e atividades e que tal publicação na web complementa o interesse e a obrigação institucional quanto à transparência de suas ações, deve ser abordado no escopo da Arquivologia, como um documento arquivístico.

Se a análise partir da função objetiva de difusão das ações e serviços institucionais, pode ser entendido como publicação oficial e interpretado como objeto da Biblioteconomia.

Finalmente, há que se observar que a preservação de *websites* pode ser feita a partir de uma visão museológica. A Lei 11904/2006, que institui o estatuto dos museus, explicitamente declara

> Consideram-se bens culturais passíveis de musealização os bens móveis e imóveis de interesse público, de natureza material ou imaterial, tomados individualmente ou em conjunto, portadores de referência ao ambiente natural, à identidade, à cultura e à memória dos diferentes grupos formadores da sociedade brasileira. (Art. 5º, §1º)

Desta forma, os websites poderiam estar contemplados no entendimento de que acervo museológico é o "conjunto formado pelos testemunhos materiais, dos mais variados suportes, formatos, materiais e origens, e imateriais dos povos e seu ambiente que são selecionados intencionalmente por seu valor de representatividade e memória" (SÃO PAULO. Secretaria de Cultura, 2014, Art. 1º, I).

A expressão chave aqui é "selecionado intencionalmente", pois indica que havendo uma decisão institucional para selecionar determinados websites como por sua representatividade quanto a um determinado evento ou tema, pode-se inseri-los numa linha de tratamento museológico.

Esse capítulo aborda aspectos legais vinculados à coleta, preservação e acesso à websites selecionados para serem mantidos como registros de memória das ações do Estado, das instituições e da sociedade, de uma forma geral. Em complemento, nesta parte introdutória, considera-se adequado tecer algumas considerações quanto aos debates que vêm ocorrendo no Congresso Nacional no que respeita a esse tema.

É uma das atribuições do Congresso Nacional, por meio de suas duas Casas, a Câmara dos Deputados e o Senado Federal, legislar sobre temas de interesse público na esfera federal. Para ilustrar essa necessidade, observa-se que a Lei nº 8.159/1991 (Lei dos Arquivos) tem proposta iniciada pelo Projeto de Lei nº 4.895/1984 da Câmara dos Deputados, por demanda do Poder Executivo, especificamente de interesse do Arquivo Nacional, órgão vinculado à época ao Ministério da Justiça. A Lei nº 10.994/2004 (Lei do Depósito Legal) surgiu de um Projeto de Lei, o PLS 110/1988, desta feita, de iniciativa do Senado Federal. Também a Lei n° 11.904/2009 (Estatuto dos Museus) foi iniciada na Câmara dos Deputados no PL 7568/2006.

No momento, a Câmara está discutindo projetos de lei ordinária que têm impacto na preservação de website. Santos (2020) fez uma análise sobre essas proposições há alguns anos, motivo pelo qual aqui serão destacados apenas dois deles, atualizando aquele estudo:

- PL 2431/2015. Dispõe sobre o patrimônio público digital institucional inserido na rede mundial de computadores e dá outras providências.
- PL 1473/2023. Esta Lei torna obrigatória a disponibilização, por parte das empresas que operam sistemas de inteligência artificial, de ferramentas que garantam aos autores de conteúdo na internet a possibilidade de restringir o uso de seus materiais pelos algoritmos de inteligência artificial, com o objetivo de preservar os direitos autorais.

Ambos os projetos de lei estão há mais de um ano sem tramitação, mas podem retomar sua discussão a qualquer momento. Assim, é preciso acompanhar os debates para adequar as políticas de preservação digital da web em conformidade com a legislação em vigor.

LEGISLAÇÃO - TERCEIRA ETAPA: Conselho Nacional de Arquivos

No que respeita à abordagem arquivística para a preservação de websites, cabe registrar, inicialmente, que o Conselho Nacional de Arquivos - Conarq é o órgão central do Sistema Nacional de Arquivos - Sinar, conforme define a Lei nº 8.159/1991 (Art. 26), e é responsável pela definição da política nacional de arquivos. Essa atribuição é formalizada por meio de resoluções.

Em 2005, o Conselho Nacional de Arquivos, explicitou em sua Carta para a Preservação do Patrimônio Arquivístico Digital, que "os documentos arquivísticos exclusivamente em formato digital, [...] como os sítios da internet, dentre muitos outros formatos e apresentações possíveis de um vasto repertório de diversidade crescente", que faziam parte do patrimônio digital (CONARQ, 2005, p. 1), ou seja, na mesma perspectiva da Carta da Unesco (2003) (TERRADA, 2022, p. 67).

Em 2023, o Conarq aprovou, no âmbito do Sinar, a Resolução nº 52, que estabelece a política de preservação de websites e mídias sociais e a Resolução nº 53, que define requisitos mínimos de preservação para websites e mídias sociais, tais requisitos devem ser em consonância com a Resolução nº13 de 9 de fevereiro de 2001, que dispõe sobre a implantação de uma política municipal de arquivos, sobre a construção de arquivos e de websites de instituições arquivísticas.

LEGISLAÇÃO - TERCEIRA ETAPA: Depósito Legal

No viés bibliográfico, projetos de preservação digital governamentais poderiam se orientar pela interpretação dos websites como publicações oficiais e, nesse sentido, estarem sujeitos à necessidade de depósito legal na Biblioteca Nacional.

Conforme a legislação específica, depósito legal corresponde à "exigência estabelecida em lei para depositar, em instituições específicas, um ou mais exemplares, de todas as publicações, produzidas por qualquer meio ou processo, para distribuição gratuita ou venda" (Lei nº 10.994/2004, Art. 2º, I).

A Lei abrange, de forma explícita, "as publicações oficiais dos níveis da administração federal, estadual e municipal, compreendendo ainda as dos órgãos e entidades de administração direta e indireta, bem como as das fundações criadas, mantidas ou subvencionadas pelo poder público" (Lei nº 10.994/2004, Art. 3º).

Neste caso, há que se observar a Lei nº 9.610/1998, que trata sobre direitos autorais e dá outras providências, pois se for considerado o website como uma publicação ou uma obra (Lei nº 9.610/1998, Art. 5º, I e VIII), deve-se levar em conta os dispositivos da referida legislação, "expressas por qualquer meio ou fixadas em qualquer suporte, tangível ou intangível" (Lei nº 9.610/1998, Art. 7º).

LEGISLAÇÃO - TERCEIRA ETAPA: Acesso a websites preservados

A legislação que regulamente o acesso à informação é ampla e diversa, por isso, no escopo deste Guia, serão discutidas apenas dois regulamentos, a lei de acesso à informação (Lei nº 11.527/2011) e a lei geral de proteção de dados pessoais (Lei nº 13.709/2018).

A Lei nº 11. 527/2011, mostra que o Estado deve garantir o direito de acesso à informação, desde que proteja "os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural" (Lei nº 13.709/2019). Deste modo, o acesso aos websites preservados por instituições como arquivos, bibliotecas e museus, deverão franquear o acesso respeitando tais normativos legais.

REFLEXÃO

A preservação de conteúdo web representa um dos desafios mais complexos e urgentes da era digital contemporânea. Ao longo deste guia, foram apresentadas diretrizes técnicas, metodológicas e legais que fundamentam as boas práticas no arquivamento da web. No entanto, é necessário refletir criticamente sobre as implicações, limitações e perspectivas futuras desta área de conhecimento em constante evolução.

A volatilidade inerente do conteúdo web constitui um paradoxo fundamental: enquanto a internet se tornou o principal repositório de conhecimento da humanidade, sua natureza dinâmica e efêmera ameaça constantemente a continuidade da memória digital. Como evidenciado pelos estudos citados neste guia, cerca de 80% das páginas web perdem sua forma original após um ano, revelando a urgência de uma resposta coordenada e sistemática. Esta reflexão nos leva a questionar não apenas como preservar, mas o que preservar. A seleção de conteúdos para arquivamento envolve inevitavelmente escolhas subjetivas que refletem valores culturais, políticos e sociais específicos. Quem decide o que merece ser preservado para as gerações futuras? Como garantir que essas decisões não reproduzam vieses ou exclusões sistemáticas?

A evolução tecnológica acelerada apresenta um duplo desafio: ao mesmo tempo que oferece novas ferramentas e possibilidades para o arquivamento da web, exige adaptação constante dos métodos e infraestruturas de preservação. A transição de páginas estáticas baseadas em HTML para aplicações complexas com JavaScript dinâmico exemplifica essa tensão entre inovação e preservação. As ferramentas apresentadas neste guia - Heritrix, Archive-It, Webrecorder, Conifer - representam respostas tecnológicas a diferentes necessidades, mas também revelam a fragmentação do campo. A ausência de uma solução única e definitiva reflete a complexidade inerente do problema, mas também pode dificultar a implementação de políticas consistentes de preservação. A padronização através do formato WARC (ISO 28500:2017) representa um avanço significativo na direção da interoperabilidade, mas sua adoção ainda não é universal. A necessidade de capacitação técnica especializada continua sendo uma barreira para muitas instituições, especialmente aquelas com recursos limitados.

A preservação de conteúdo web levanta questões éticas complexas sobre privacidade, direitos autorais e acesso à informação. O arquivamento de sites pode envolver a captura inadvertida de dados pessoais, comentários de usuários ou conteúdos protegidos por direitos autorais. Como equilibrar o interesse público na preservação da memória digital com os direitos individuais à privacidade e propriedade intelectual? A legislação brasileira, ainda em desenvolvimento nesta área, reflete a tensão entre diferentes marcos regulatórios. A Lei Geral de Proteção de Dados (LGPD) e a Lei de Acesso à Informação estabelecem princípios por vezes contraditórios que requerem interpretação cuidadosa no contexto do arquivamento web. As proposições legislativas em tramitação no Congresso Nacional, como os PLs 2431/2015 e 1473/2023, indicam a necessidade de um debate mais amplo sobre o tema.

A análise do panorama internacional revela disparidades significativas na capacidade de preservação digital entre diferentes países e regiões. Enquanto nações desenvolvidas possuem programas consolidados de arquivamento web, muitos países em desenvolvimento, incluindo o Brasil, ainda carecem de iniciativas estruturadas. Esta desigualdade tem implicações profundas para a representação global da memória digital. Se apenas alguns países preservam sistematicamente seu patrimônio web, corremos o risco de criar uma memória digital fragmentada e enviesada, que reflete principalmente as perspectivas dos países com maior capacidade técnica e financeira. O caso do Chile, único país latino-americano com programa consolidado no IIPC, ilustra tanto as possibilidades quanto os desafios regionais. A colaboração internacional, embora essencial, não pode substituir o desenvolvimento de capacidades locais e políticas nacionais apropriadas.

A preservação digital é um compromisso de longo prazo que requer investimento contínuo em infraestrutura, tecnologia e recursos humanos. A sustentabilidade financeira dos projetos de arquivamento web permanece um desafio crítico, especialmente considerando o crescimento exponencial do volume de dados a serem preservados. O modelo de colaboração internacional representado pelo IIPC oferece uma alternativa importante, permitindo o compartilhamento de custos e expertise. No entanto, a dependência de organizações estrangeiras pode criar vulnerabilidades estratégicas e limitar a autonomia nacional na definição de políticas de preservação.

As tecnologias emergentes, como inteligência artificial e blockchain, apresentam tanto oportunidades quanto desafios para o arquivamento web. A IA pode automatizar e otimizar processos de seleção e catalogação, mas também levanta questões sobre transparência e controle humano nas decisões de preservação. O blockchain oferece possibilidades interessantes para garantia de integridade e proveniência, mas sua sustentabilidade energética permanece questionável. A crescente importância das redes sociais e plataformas digitais como espaços de produção cultural e debate público torna ainda mais urgente o desenvolvimento de estratégias específicas para esses ambientes. As tradicionais ferramentas de crawling mostram-se inadequadas para capturar a complexidade das interações sociais digitais.

A complexidade técnica e conceitual do arquivamento web exige profissionais qualificados capazes de navegar entre diferentes disciplinas - arquivologia, biblioteconomia, ciência da computação, direito. A formação desses profissionais requer currículos atualizados e abordagens interdisciplinares que ainda não são amplamente disponíveis. A colaboração entre universidades, instituições de memória e organizações técnicas é essencial para desenvolver competências adequadas. A experiência internacional mostra que o sucesso dos programas de arquivamento web depende fundamentalmente da qualidade das equipes envolvidas.

Um aspecto frequentemente negligenciado é a dimensão participativa do arquivamento web. Como envolver comunidades locais, grupos marginalizados e cidadãos comuns nos processos de seleção e preservação? Como garantir que a memória digital preserve não apenas os discursos oficiais, mas também as vozes diversas da sociedade? Ferramentas como o Conifer democratizam parcialmente o arquivamento web ao permitir que indivíduos capturem conteúdos sem conhecimento técnico avançado. No entanto, a preservação de longo prazo ainda requer infraestruturas institucionais robustas.

A preservação de conteúdo web não é apenas uma questão técnica, mas um imperativo democrático e cultural. Em uma sociedade cada vez mais dependente da informação digital, a capacidade de manter acessível nossa memória coletiva torna-se fundamental para a continuidade cultural e o exercício da cidadania. Os desafios são significativos: a volatilidade dos conteúdos, a evolução tecnológica acelerada, as limitações de recursos, as questões legais e éticas, as disparidades globais. No entanto, os avanços tecnológicos, a crescente consciência sobre a importância do tema e as iniciativas de colaboração internacional oferecem razões para otimismo.

As iniciativas pioneeiras como o ARQUIWEB do IBICT e as resoluções do CONARQ estabelecem fundamentos importantes, mas é necessário um esforço coordenado para desenvolver uma política nacional abrangente de preservação digital da web. Este guia representa uma contribuição para esse esforço coletivo, mas sua efetividade dependerá da apropriação e adaptação por parte de instituições, profissionais e formuladores de políticas públicas. A preservação da memória digital é responsabilidade compartilhada que requer engajamento contínuo de toda a sociedade. A web que conhecemos hoje será história amanhã. Nossa responsabilidade é garantir que essa história permaneça acessível, compreensível e útil para as gerações futuras.

REFERÊNCIAS

ARELLANO, Miguel Angel. Preservação de documentos digitais. Ciência da informação, v. 33, p. 15-27, 2004. Disponível em: https://www.scielo.br/j/ci/a/FLfgJvpH3PZKf3HbpKYchZr/?format=pdf&lang=pt. Acesso em: 11 nov. 2023

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR ISO 14721: sistemas espaciais de transferência e de informação - Sistema Aberto de Arquivamento de Informação (SAAI) - Modelo de referência. Rio de Janeiro: ABNT, 2021. 121p.

BIBLIOTHÈQUE NATIONALE DU LUXEMBOURG. Rapport d'activité 2022. Luxembourg: Bibliothèque nationale du Luxembourg, 2023. Disponível em: https://bnl.public.lu/fr/a-la-une/publications/rapports-annuels/rapport-activite-2022.html. Acesso em: 11 nov. 2024.

BOERES, Sonia Araújo de Assis; SAAD, Rondineli Gama. Arquivamento da Web: definições, estratégias, fluxos e iniciativas. Revista Brasileira de Preservação Digital, Campinas, SP, v. 4, n. 00, p. e023005, 2023. DOI: 10.20396/rebpred.v4i00.17934. Disponível em: https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/17934. Acesso em: 11 nov. 2024.

BOERES, Sonia Araújo de Assis. Competências necessárias para equipes de profissionais de preservação digital. 2017. Disponível em: http://www.rlbea.unb.br/jspui/bitstream/10482/24354/1/2017_SoniaAraujodeAssisBoeres.pdf. Acesso em: 11 nov. 2024

BRASIL. Lei nº 10.994, de 14 de dezembro de 2004.Dispõe sobre o depósito legal de publicações, na Biblioteca Nacional, e dá outras providências. Diário Oficial União: Brasília, DF,14 de dezembro de 2004. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/l10994.htm. Acesso em: 31 out. 2024.

BRASIL. Lei nº 9.610, de 19 de fevereiro de 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências. Diário Oficial União: seção 1, Brasília, DF,19 de fevereiro de 1998. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9610.htm. Acesso em: 31 out. 2024.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º , no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências.Diário Oficial União: Brasília, DF,18 de novembro de 2011. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm. Acesso em: 31 out. 2024.

BRASIL. Lei nº 13.709, DE 8 de Julho DE 2019. Altera a Lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados; e dá outras providências. Diário Oficial União: Brasília, DF, 19 de dezembro de 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Lei/L13853.htm#art1. Acesso em: 31 out. 2024

BRÜGGER, Niel. Archiving Websites: General Considerations and Strategies. Århus: The Centre for Internet Research, 2005.

BRÜGGER, Niels; FINNEMANN, Niels Ole. The web and digital humanities: theoretical and methodological concerns. Journal of Broadcasting and Electronic Media, v. 57, n.1, 2013.

BRITISH LIBRARY. Legal deposit and web archiving.

BROWN, Adrian. Archiving Websites: a practical guide for information management professionals. Facet publishing: London, 2006.

CONSELHO NACIONAL DE ARQUIVOS. Resolução nº 13, de 09 de fevereiro de 2001. Dispõe sobre a implantação de uma política municipal de arquivos, sobre a construção de arquivos e de websites de instituições arquivísticas. Disponível em: https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-13-de-9-de-fevereiro-de-2001. Acesso em: 31 out. 2024.

CONSELHO NACIONAL DE ARQUIVOS – CONARQ. Câmara Técnica de Documentos Eletrônicos. Carta para a preservação do patrimônio arquivístico digital. [S. l.]: CONARQ, 2005. Disponível em: http://conarq.gov.br/images/publicacoes_textos/Carta_preservacao.pdf. Acesso em: 31 de maio de 2021.

CONSELHO NACIONAL DE ARQUIVOS. Resolução nº 52, de 25 de agosto de 2023. Estabelece a política de preservação de websites e mídias sociais no âmbito do Sistema Nacional de Arquivos (Sinar). Disponível em: https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-52-de-25-de-agosto-de-2023 Acesso em: 17 out. 2024.

CONSELHO NACIONAL DE ARQUIVOS.Resolução nº 53, de 25 de agosto de 2023. Define requisitos míniomos de preservação de websites e mídias sociais no âmbito do Sistema Nacional de Arquivos (Sinar). Disponível em: https://www.gov.br/conarq/pt-br/legislacao-arquivistica/resolucoes-do-conarq/resolucao-no-53-de-25-de-agosto-de-2023 Acesso em: 17 out. 2024.

COSTA, Miguel; GOMES, Daniel; SILVA, Mário J. The evolution of web archiving. International Journal on Digital Libraries, 1-15. doi: 10.1007%2Fs00799-016-0171-9, 2017.

COSTA, Miguel. Information search in web archives. Disponível em: https://repositorio.ul.pt/bitstream/10451/16020/1/ulsd069905_td_Miguel_Costa.pdf. Acesso em: 12 jul 2020.

COUNCIL OF THE CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS (CCSDS). Reference Model for an Open Archival Information System (OAIS): Recommended Practice, Issue 2, CCSDS, 652.0-M-2. Magenta Book. Washington, DC, USA . Council of the Consultative Committee for Space Data Systems, June 2012. 135 p. Disponível em: https://public.ccsds.org/pubs/650x0m2.pdf . Acesso em: 12 set. 2024.

FERREIRA, Lisiane Braga; MARTINS, Marina Rodrigues; ROCKEMBACH, Moisés. Usos do arquivamento da web na comunicação científica, 2018. DOI 10.21747/16463153. Disponível em: <http://ojs.letras.up.pt/index.php/prismacom/article/view/3927>. Acesso em: 28 mai. 2018.

FERREIRA, Lisiane Braga. Arquivamento da web e mídias sociais : preservação digital de vídeos da campanha presidencial brasileira de 2018. Dissertação. 2019. Mestrado em comunicação e Informação. Universidade Federal do Rio Grande do Sul. Disponível em: http://hdl.handle.net/10183/194617

FORMENTON, Danilo; GRACIOSO, Luciana de Souza. Padrões de metadados no arquivamento da web: recursos tecnológicos para a garantia da preservação digital de websites arquivados. RDBCI: Rev. Dig. Bibliotec e Ci. Info., v. 20, 2022. Disponível em: https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8666263/27829. Acesso em: 28 set. 2022.

GOMES, Daniel. Preservar a web: um desafio ao alcance de todos. In: Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas, 10, 2010, Guimarães. Actas... Lisboa: B.A.D., 2010. (on-line)

GOMES, Daniel.; MIRANDA, João; COSTA, Miguel. A survey on web archiving initiatives. In: International Conference on Theory and Practice of Digital Libraries. Lisboa, POR: Springer Berlin Heidelberg, 2011, p. 408-420. Disponível em: <https://link.springer.com/content/pdf/10.1007%2F978-3-642-24469-8_41.pdf>. Acesso em: 06 out. 2017.

HARVARD UNIVERSITY. Web Archives Collections. 2017. Disponível em: https://preservation.library.harvard.edu/web-archives-collections. Acesso em: 11 nov. 2024.

HOCKX-YU, H. The past issue of the web. In: WEBSCI 11 PROCEEDINGS OF THE 3RD INTERNATIONAL WEB SCIENCE CONFERENCE, n 12, Jun. 2011, New York, NY. Proceedings [...]. New York, NY, USA: Association for Computing Machinery, 2011. p. 1–8. Disponível em: https://doi.org/10.1145/2527031.2527050. Acesso em: 21 abr. 2021.

IBICT. ARQWEB, 2022. Disponível em: http://arqweb.ibict.br/ . Acesso em: 20 abr. 2023.

LIBRARY OF CONGRESS. Preservation: recommended formats statement - web archives. Disponível em: https://www.loc.gov/preservation/resources/rfs/webarchives.html Acesso em: 10 ago. 2021.

LIBRARY OF CONGRESS. Library of Congress Collections Policy Statements. Washington, D.C.: Library of Congress, 2022. Disponível em: https://www.loc.gov/acq/devpol/. Acesso em: 11 nov. 2024.

LOHNDORF, Anja. Web Archiving at the National and University Library of Iceland. Alexandria: The Journal of National and International Library and Information Issues, v. 24, n. 1, p. 1-10, 2013.

LUZ, Ana Javes. Preservação de sites oficiais: exemplos internacionais e o caso brasileiro. Revista Brasileira de Preservação Digital, Campinas, SP, v.3, 2022. DOI: 10.20396/rebpred.v3i00.16587. Acesso em: 18 maio 2023.

KULOVITS, Hannes. Plato: a preservation planning tool. 2009. Disponível em: https://typeset.io/papers/plato-a-preservation-planning-tool-31qarzyn83. Acesso em: 11 nov. 2024.

MAEMURA, Emily et al. If these crawls could talk: Studying and documenting web archives provenance. Journal of the Association for Information Science and Technology, v. 69, n. 10, p. 1223–1233, 2018.

MASANÉS, Julien. Selection for web Archives. In: MASANÉS, Julien. Web Archiving. Berlin: Springer, Heidelberg, 2006. p. 71-90.

MARTINS, Marina Rodrigues; ROCKEMBACH, Moises. Promoção de iniciativas de arquivamento da web: um estudo a partir da rede de públicos estratégicos da UFRGS. AtoZ: novas práticas em informação e conhecimento. Curitiba: Universidade Federal do Paraná, Programa de Pós-Graduação em Gestão da Informação. Vol. 8, n. 2 (jul./dez. 2019), p. 99-105, 2019

MELO, J. F. Arquivamento dos websites do governo federal brasileiro: preservação do domínio gov.br. 2020. 133 f. Dissertação (Mestrado – Programa
de Pós-Graduação em Comunicação e Informação) – Universidade Federal do Rio
Grande do Sul, Porto Alegre, 2020. Disponível em: http://hdl.handle.net/10183/210671 . Acesso em: 04 abr. 2023.

PENNOCK, Maureen. Web-Archiving. York: Digital Preservation Coalition, 2013. (DPC Technology Watch Report 13-01). Disponível em: https://www.dpconline.org/docs/dpc-technology-watch-publications/technology-watch-reports-1/865-dpctw13-01-pdf/file. Acesso em: 11 nov. 2024.

ROCKEMBACH, Moisés; PAVÃO, Caterina Marta Groposo. Políticas e tecnologias de preservação digital no arquivamento da Web. Revista Ibero-Americana de Ciência da Informação (RICI). Brasília, v. 11, n. 1, 2018. Disponível em: https://lume.ufrgs.br/handle/10183/175153

ROCKEMBACH, Moisés; PAVÃO, Caterina Marta Groposo. Arquivamento da web e preservação digital. São Paulo: Pimenta Cultural, 2024. Disponível em: https://www.pimentacultural.com/livro/arquivamento-web/ . Acesso em: 12 set. 2024.

ROCKEMBACH, Moisés. Arquivamento da Web: estudos de caso internacionais e o caso brasileiro. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, v.16, n.1, p. 7-24, 2018. https://doi.org/10.20396/rdbci.v16i1.8648747

SANTOS, Vanderlei Batista dos. Arquivamento web : legislação correlata. Revista Brasileira de Preservação Digital, Campinas, SP, v. 1, n. 00, p. e020005, 2020. Disponível em: https://econtents.bc.unicamp.br/inpec/index.php/rebpred/article/view/14800. Acesso em: 10 ago. 2021.

SÃO PAULO. Secretaria de Cultura. Resolução SC 105/2014. Estabelece princípios, procedimentos e fixa normas para recebimento e incorporação de bens móveis que constituem acervos museológicos, arquivísticos e documentais e de obras raras de natureza bibliográfica, pelas modalidades de doação, legado, coleta, permuta, transferência definitiva sem encargos e compra, pelos museus da Secretaria de Estado da Cultura. Disponível em: https://www.imprensaoficial.com.br/DO/GatewayPDF.aspx?link=/2014/executivo%20secao%20i/novembro/12/pag_0043_4S2SK0B6SRFG3e85OPHOL6J435L.pdf Acesso em: 17 out. 2024.)

SILVA, Karina Moura da. Um modelo de ciclo de vida de dados na web. Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife. Orientadora: Bernadette Farias Lóscio. 107 f.: il., fig., tab. 2019. Disponível em: < https://repositorio.ufpe.br/handle/123456789/34147 >. Acesso em 28 out.2021.

TERRADA, Gabriela Ayres Ferreira. Preservação digital da web: uma reflexão sobre políticas e práticas. 2022. 215 f. Dissertação (Mestrado em Ciência da Informação) - Instituto de Arte e Comunicação Social, Universidade Federal Fluminense, Niterói, 2022. Disponível em: https://app.uff.br/riuff/handle/1/26276 . Acesso em: 31 out. 2024.

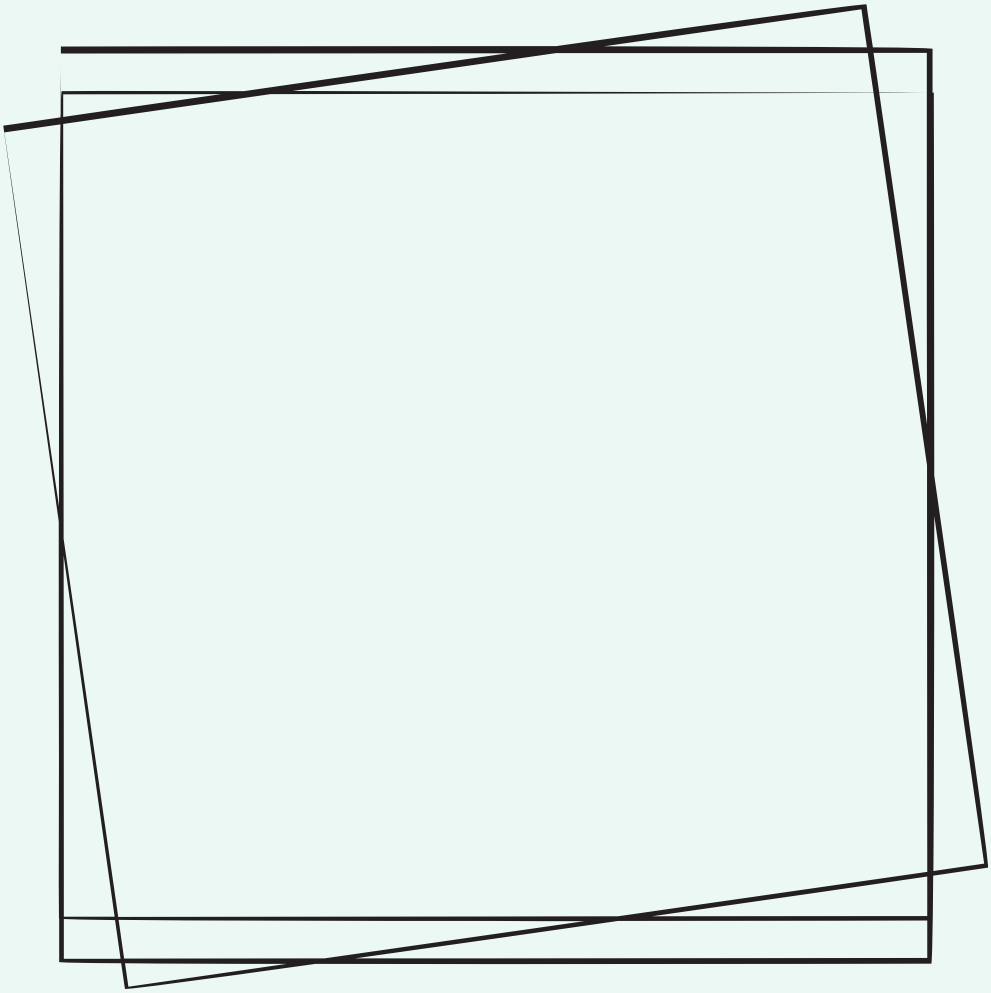
UNESCO. Carta sobre la preservación del patrimonio digital, 15 out. 2003.

VENLET, Jessica et al. Descriptive metadata for web archiving: literature review of user needs. Dublin, Ohio: Online Computer Library Center (OCLC) Research, Feb. c2018. 48 p. Disponível em: https://www.oclc.org/content/dam/research/publications/2018/oclcresearch- wam-literature-review-user-needs.pdf. Acesso em: 7 jun. 2023

VLASSENROOT, Eveline et al. Web-archiving and social media: an exploratory analysis: Call for papers digital humanities and web archives–A special issue of international journal of digital humanities. International Journal of Digital Humanities, v. 2, n. 1-3, 107-128, 2021. DOI 10.1007/s42803-021-00036-1

GLOSSÁRIO

https://glossario.cariniana.ibict.br/vocab/index.php



Dríade